University of Basel
Faculty of Business and Economics

WWZ

March 2020

Estimating the Green Potential of Occupations: A New Approach Applied to the U.S. Labor Market

WWZ Working Paper 2020/03

Christian Rutzer, Matthias Niggli, Rolf Weder

A publication of the Center of Business and Economics (WWZ), University of Basel.

Universität Basel
Peter Merian-Weg 6
4052 Basel, Switzerland
wwz.unibas.ch

Corresponding Author:
Christian Rutzer
christian.rutzer@unibas.ch

Estimating the Green Potential of Occupations: A New Approach Applied to the U.S. Labor Market

Christian Rutzer, Matthias Niggli and Rolf Weder*

March 2020

Abstract

This paper presents a new approach to estimate the green potential of occupations. Using data from O*NET on the skills that workers possess and the tasks they carry out, we train several machine learning algorithms to predict the green potential of U.S. occupations classified according to the 6-digit Standard Occupational Classification. Our methodology allows existing discrete classifications of occupations to be extended to a continuum of classes. This improves the analysis of heterogeneous occupations in terms of their green potential. Our approach makes two contributions to the literature. First, as it more accurately ranks occupations in terms of their green potential, it leads to a better understanding of the extent to which a given workforce is prepared to cope with a transition to a green economy. Second, it allows for a more accurate analysis of differences between workforces across regions. We use U.S. occupational employment data to highlight both aspects.

Keywords: green skills, green tasks, green potential, supervised learning, labor market

JEL codes: C53, J21, J24, Q52

*Faculty of Business and Economics, University of Basel. Correspondence address: Christian Rutzer, Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, CH-4002 Basel, Switzerland, email: christian.rutzer@unibas.ch. This work has been supported by the Swiss National Science Foundation (SNSF) within the framework of the National Research Program "Sustainable Economy: resource-friendly, future-oriented, innovative" (NRP 73), Grant-№ 407340-172430.

1 Introduction

Unambiguous signs of a global climate change emphasized by natural scientists, huge wildfires in Western Canada, rapidly melting glaciers in Switzerland and polar bears threatened by the Arctic ice meltdown trigger an increasing demand for a *green transition* of economies. The question, however, is how this transition is managed in an efficient way and what effects this may have on societies, economies and the environment. An important concern is that the acceptance of a green transition may largely depend on how individual workers are able to cope with the implied new requirements in the labor market (Martinez-Fernandez et al., 2010; Bowen et al., 2016). The success of the transition may, in turn, depend on the availability of workers who can offer the required *green skills*. Thus, from a policy perspective, it is important to know (i) which occupations could benefit from a green transition and (ii) how large the share of the workforce is that belongs to these occupations. In this paper, we address both questions using a novel approach based on machine learning techniques.

We build on important contributions in the literature that determine which workers may benefit from a green transition. Using discrete occupation classifications, these approaches typically define groups of occupations (e.g. "green jobs", "non-green jobs") and compare them in terms of, e.g., employment shares, average education levels or routine-task intensities. Consoli et al. (2016) use current green occupations in the United States as defined by O*NET[1] and find that a significant share of today's workers in the U.S. could be facing changing job characteristics due to the structural change induced by greening the economy (i.e., 9.8% - 12.3% of aggregate employment in 2011-2012). They emphasize that, compared to similar non-green occupations, these "green jobs" seem to require more abstract skills and conduct fewer routine tasks. With this contribution, Consoli et al. (2016) have presented a comprehensive overview of green employment that describes well the basic characteristics of green occupations.

Bowen et al. (2018), however, emphasize that a binary classification of occupations cannot account for the fact that there are non-green occupations which nevertheless are "potentially" green. These jobs, they argue, are to some degree similar to green jobs. In these cases, the job holders could, in principle, also perform green tasks and, consequently, adjust to the transition to green occupations rather easily and even benefit from a green transition. Bowen et al. (2018) call these occupations "green rival jobs" and define them as those occupations for which O*NET

[1]see https://www.onetonline.org/find/green/ for more information on the data they use.

states that they have at least one green job among their registered "similar jobs". With this approach, they extend the classification of occupations based on Consoli et al. (2016) to three major classes of occupations: "green occupations", "green rival occupations" and "other (non-green) occupations".[2] Based on U.S. employment data from 2014, Bowen et al. (2018) estimate that "at the national level, 19.4 percent of workers could currently have green jobs. This group consists of "directly green" occupations (10.3%), which corresponds to the green jobs considered by Consoli et al. (2016), and, additionally "indirectly green" occupations (9.1%). Moreover, 44.3 percent are in green rival jobs, and 36.3 percent have other jobs" (p. 265).

Bowen et al. (2018) also discuss some limitations of their approach to capture potentially green jobs. First, as they use the category "similar jobs" from O*NET as an indicator for the green potential of an occupation, they cannot directly compare what matters most to perform green tasks, i.e., specific skill sets as emphasized by (Vona et al., 2018). Hence, the approach suggested by Bowen et al. (2018) may identify green rival jobs because of some similarities with green jobs that do not matter in order to perform green tasks. Second, there is no additional information on *how similar* specific green rival jobs are compared to green jobs. This degree of similarity is relevant because, as Bowen et al. (2018) point out themselves, there is substantial heterogeneity within the identified occupational groups (e.g., in terms of the skill content or of the relative number of green tasks). Some green rival jobs may thus be very similar to corresponding green jobs whereas others may be rather different. This aspect cannot be addressed with their approach. Consequently, they argue that instead of focusing on discrete green occupation-classifications (e.g. green jobs, non-green jobs, green rival jobs), future research should assess *how green* occupations are (i.e. to consider the green potential of an occupation as a continuous variable).

This is where we intend to make a contribution with this paper. We develop an approach that allows to estimate the *green potential* of occupations in the U.S. more accurately than preceding techniques. Following Vona et al. (2018), we define the green potential of an occupation based on the skills which are required to perform green tasks. The basic intuition of this approach is simple: There are some skills which are especially relevant (or irrelevant) to perform green tasks. And the more (fewer) of these skills an occupation requires, the higher (lower) its green potential. In

[2] In fact, Bowen et al. (2018) analyze five different groups of occupations. As Consoli et al. (2016), they start with three groups defined by O*NET ("green enhanced skills", "green new and emerging" and "green increased demand") and split the remaining jobs into "green rival jobs" and "other jobs". Thereby, they consider "green enhanced skills" and "green new and emerging" jobs as "directly green" occupations, "green increased demand" as "indirectly green", "rival green" as potentially green and "other jobs" as non-green.

other words, a high green potential means that such an occupation basically requires the skills and competences which are necessary to fulfill green tasks. In O*NET, two examples of green tasks would be to "prepare, review, or update environmental investigation or recommendation reports" or "identify and recommend energy savings strategies to achieve more energy-efficient operations". However, whether an occupation currently performs green tasks or not is irrelevant.

Our approach is thus similar to Vona et al. (2018), who identify a specific set of skills that are especially important to perform green tasks. In order to find these skills, they regress, at the 8-digit O*NET occupation level, the relative number of green tasks on each general skill provided by O*NET. Subsequently, they classify each skill with a positive and, at the 99% level statistically significant, coefficient as an indicator that captures the importance of this skill to perform green tasks. This delivers a set of skills (e.g. "Biology"), which are important to perform green tasks. Using principal component analysis (PCA), they group these skills to four so-called "Green General Skills" (GGS) indices. For example, their GGS "Science" consists of the skills "Physics" and "Biology", which have both been identified to be important for green tasks by OLS. For a given occupation the GGS scores are thus measures of the importance of a specific GGS in that occupation. In other words, the approach of Vona et al. (2018) allows us to evaluate the importance of selected green skills not only in green but also in non-green occupations, which can be interpreted as an occupation's green potential.[3]

Their technical approach is somewhat different to ours. By using GGS scores, they use only an unweighted average of a subset of skills to determine the potential of an occupation to perform green tasks. In contrast to that, we want to determine the green potential of occupations based on their *entire* skill set. In order to do so, we use a slightly different estimation approach. We also use skills-data from O*NET, but differently to Vona et al. (2018), we train several machine learning algorithms to predict the green potential of occupations. These algorithms allow us to, first, investigate the full skill set of occupations and, second, to find out the optimal weighting structure of skills to predict the green potential of occupations. Based on commonly used statistical tests, all of our trained machine learning algorithms perform significantly better than the GGS scores proposed by Vona et al. (2018) to predict the green potential of occupations.

Taken together, our approach makes the following contribution to the literature. We develop

[3]It is important to note that, in contrast to Bowen et al. (2018), Consoli et al. (2016) and our paper, the work of Vona et al. (2018) has a different research question. Their aim is not to determine the green potential of a country's workforce, but rather to analyze whether occupations with many "green skills" are differently affected by environmental regulations than those with few of these skills.

a green potential measure that is continuous and based on the entire skill sets of occupations. This leads to more accurate predictions of the green potential of jobs from a statistical point of view. Thus, it allows us to rank occupations more accurately in terms of their green potential than previous approaches. Our analysis implies that occupations which require a relatively large number of technical skills also have the highest green potential. Social, cultural and artistic occupations, in turn, tend to have low green potential.

On the aggregate level, our approach leads to a more nuanced characterization of potentially green employment, as it allows us to investigate the occupation and employment distribution in much more detail compared to existing work. We illustrate this claim by applying our approach to the U.S. labor market. We show that our approach results in a more accurate grouping of occupations on the federal level and makes it possible to highlight differences across states that could not be discovered when using existing discrete approaches.

The remainder of the paper is structured as follows. In Section 2, we carefully explain our identification strategy to capture the green potential of occupations and describe the data used to predict the green potential of occupations. In Section 3, we train different machine learning algorithms by using the data. Section 4 discusses our estimates of the green potential at the occupation level and relates them to findings of the previous literature. In Section 5, we use our green potential estimates to assess the green potential of the U.S. labor market and discuss how our approach contributes to current research in the field. Section 6 summarizes and concludes the paper.

2 Identification and Data

In this section, we set out what we mean by the "green potential of an occupation", discuss the information which is necessary to determine it and present the data that allows us to extract this required information. We follow Autor (2013, p. 186) and define a task as a "unit of work activity that produces output" and a skill as a "worker's stock of capabilities for performing various tasks"; this implies that "[W]orkers apply their skills to tasks in exchange for wages". As emphasized by Autor (2013) the distinction between skills and tasks becomes particularly important "when the assignment of skills to tasks is subject to change" (p. 186). An increase in environmental regulations may trigger such a change as the set of tasks demanded in an economy may shift towards "green tasks". The question then arises as to what extent the workers' set of skills or capabilities is sufficient to accomplish these new tasks. This is the focus of our analysis.

Our empirical approach builds on two important insights from previous studies, analyzing green skills and green jobs, which we briefly reviewed in the Introduction. First, what matters for measuring the green potential of occupations are the tasks workers perform and not the industries they are employed in. Bowen et al. (2018) illustrate this argument with a helpful example: Consider a secretary who is employed in the renewable energy sector. Industry-based approaches would always classify this secretary's job as a "green job" (or–in our case–a "job with high green potential"). However, the position of the very same secretary would be considered a non-green job if his employer was located in the non-renewable energy sector–even though the tasks carried out and the skills required are essentially the same. Thus, this example shows that approaches building on industry classifications are of limited usefulness to assess the "green potential" of existing occupations. Therefore, our analysis will focus on the tasks carried out in different occupations.

Second, what matters for performing green tasks, are the skills that people possess and not the occupation they currently have (Vona et al., 2018). An example can again help to illustrate this claim. Consider two engineers with the same kind of education and training. One of them is involved in developing environmentally sustainable production processes, while the other is concerned with extracting natural resources. One may argue that the first engineer performs green tasks, while the second engineer does not. However, as mentioned above, the two have a similar education and, therefore, similar skills. This means that, after some training, they would be able to perform each other's tasks interchangeably. In other words, even though the resource-extracting engineer currently does not accomplish green tasks, they nevertheless have, due to their skills, a "high green potential" because they are, in principle, able to perform green tasks. Hence, what determines the green potential of an occupation is the set of skills and not whether workers are carrying out green tasks at their current workplace.

Thus, in order to determine the green potential of occupations the following information is crucial: (i) green tasks carried out in occupations, (ii) the skill sets people typically have in different occupations and (iii) a mechanism that brings points (i) and (ii) together in order to identify the necessary skills to perform green tasks. Fortunately, (i) and (ii) can be identified by using data from O*NET. We argue that (iii) can be derived by using (i) and (ii) as input to train machine learning algorithms.

Let us briefly describe the O*NET database to provide a better understanding of what kind of data we rely on to determine the green potential of occupations. O*NET is a database for occupation-level information on 965 different occupations (Version 22.0). Occupations are classified according to the US 8-digit Standard Occupational Classification (SOC) System, which is the coding system used by the U.S. government to classify occupations. Information on different occupations provided by O*NET has been used by various researchers, including the literature on green jobs (see Consoli et al., 2016; Bowen et al., 2018; Vona et al., 2018).[4]

Two pieces of information from O*NET are of particular importance to determine the green potential of occupations. With respect to the first requirement (i), O*NET contains a list of work activities for each of the 965 occupations (e.g. "performing administrative activities") and job-specific tasks (e.g. "buy, sell, or trade carbon emissions permits") which are performed by persons employed in that particular occupation. O*NET also defines whether a work activity or job-specific task is a so-called "green task". Hence, it is possible to state for each occupation the number of green tasks, T_G, relative to the number of all tasks that also include the non-green tasks, T_N. Thus, $T_G/(T_G + T_N)$.[5]

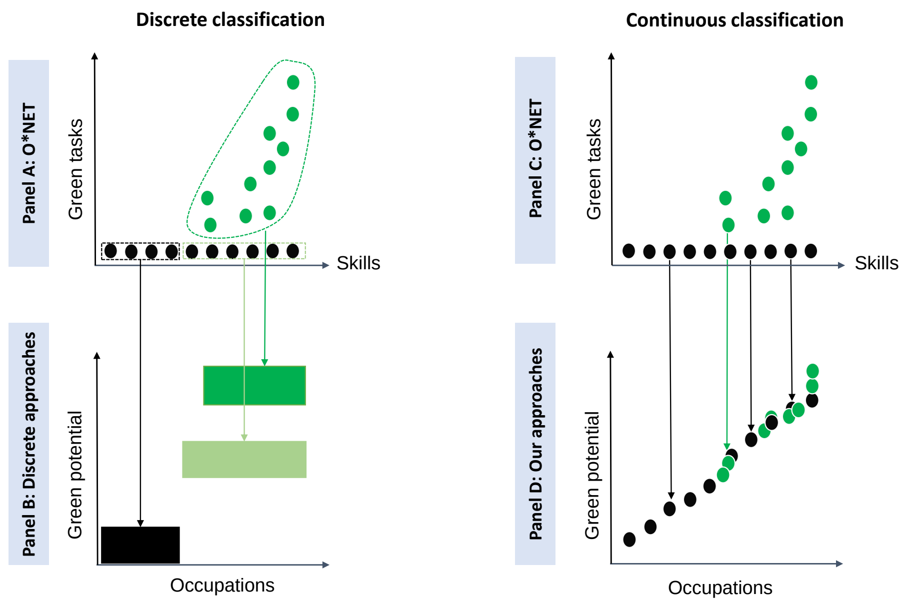
Considering the second requirement (ii), O*NET contains a fixed list of 114 rather general skills for every one of the 965 occupations, each of which is assessed regarding its "importance" to accomplish all the tasks associated with a certain occupation and its "level" (e.g., the level of mathematical expertise). This allows us to compare different occupations with regard to their skill sets. However, as the list of skills and their corresponding values is fixed, we do not have any information on the evolution of the skill sets of occupations. Thus, it is, for example, not possible to observe whether the green transition requires new skills, which could be the case for some occupations.[6] This may be the case for some occupations–particularly for occupations with rather specific skills (e.g. Strietska-Ilina et al., 2012). In contrast, the skills used in O*NET are rather general (e.g. "Complex Problem Solving" or "Mathematics") and cover a very wide range of skills (e.g. from "Time Management" to "Writing" and "Provide Consultation and Advice to Others" all the way to "Judgment and Decision-making"). Thus, the static nature of skill sets should not provide a major limitation for our analysis as the rather general skills from O*NET are unlikely to change considerably over the period we have data for. Nevertheless, we should

[4] For more information on O*NET see www.onetonline.org or Dierdorff et al. (2009).

[5] It has to be noted that O*NET allows only differentiation between green and non-green tasks.

[6] Our approach has this limitation in common with previous studies in the field (see e.g. Vona et al., 2018; Bowen et al., 2018; Consoli et al., 2016).

Figure 1: Evaluating Green Potential: Mapping Skills and Tasks



interpret the skill sets from O*NET as a *precondition* to work in a given occupation rather than as a detailed skill requirement. If, for example, the green transition requires some occupations to increasingly work with specialized software, this should not substantially change their broader skill requirements (e.g., in terms of "Complex Problem Solving" or "Mathematics" which are taken into account by O*NET). Overall, we therefore consider the data provided by O*NET well suited to investigate the green potential of occupations.

The major challenge is (iii) to empirically relate tasks, skills and occupations. In principle, such an analysis can be conducted by using a discrete or a continuous characterization of the green potential. Note, however, that the process of mapping tasks, occupations and skills is inherently different between the discrete and the continuous approach. Let us illustrate this with the help of Figure 1. Panels A and C at the top of Figure 1 both depict the information that O*NET provides regarding tasks and skills for a single occupation: Occupations are ordered with respect to both the associated tasks and skills. Further, occupations that perform a non-zero number of green tasks (e.g., measured by $T_G/(T_G + T_N)$) are illustrated with a green dot on the right of panels A and C and are associated with particular sets of skills. In addition, occupations which require the same skill set (and thus have the same value on the skill axis) may differ with respect to the relative number of green tasks. The mentioned example of the two engineers–one developing sustainable production processes and the other one extracting natural resources–may come to mind.

Let us now show the methodological difference in relating the information provided by panels A and C to occupations for a discrete and a continuous classification of occupations. The left side of Figure 1 illustrates how discrete approaches make use of the data. In short, such approaches group occupations according to common characteristics. For example, occupations with a positive number of green tasks (represented by a green dot in Panel A) can be grouped to a class of "green jobs"; these are represented by the upper green bar in Panel B. Additionally, occupations that share similarities to those in the class of "green jobs" (due to their required skills) can be grouped in another class of jobs, exemplified with the light green bar in Panel B. In contrast, the right side of Figure 1 presents how green potential can be determined using a continuous characterization. Occupations are no longer grouped based on a single common feature (e.g. non-zero green tasks). Instead, they are ranked according to the full set of skills that they possess. As a result, black-colored occupations currently not performing any green tasks can still be ranked very high in terms of green potential as shown in panel D. This is the case because–as we have argued above–the skill set associated with an occupation is the major determinant of an occupation's potential to perform green tasks.

The second approach (i.e., the continuous classification) is the direction we are going to follow in this paper. This allows us to eliminate or, at least, reduce two limitations of the discrete classification approach. First, discrete approaches neglect differences among occupations with respect to their potential for performing green tasks when occupations are grouped together as shown in panels A and B. In other words, all occupations that belong to a given group are assumed to have identical green potential (e.g., all occupations classified in the light green bar of Panel B). However, as can easily be seen by the width of the bars, there is significant heterogeneity in terms of skills. What this means is that the skills distance, and, thus the potential to perform green tasks between occupations within the same group can be substantial. This is typically not addressed with a discrete approach, but will be taken into account by the continuous classification (see Panel D). Second, the skill sets of some occupations may be very close to occupations belonging to another group. In fact, in our graphical example, the bars are even overlapping. Again, consider the two green bars in Panel B. According to O*NET information in Panel A, several occupations belonging to either of the two groups have very similar skill sets. However, they are grouped into different classes. Hence, this may understate the green potential of some occupations while it can also overstate the green potential of others. Again, a continuous characterization does not have this shortcoming.

Having recognized the advantages of the continuous approach, the question arises as to how to relate skills of occupations to tasks they can perform. In principle, several empirical techniques are possible. In the next section, we present some of the possibilities and argue that using machine learning algorithms constitutes the most promising approach.

3 Relating Skills to Green Tasks

As argued in Section 2, the green potential should depend heavily on the skill set of an occupation (and not so much on whether persons belonging to a particular occupation are currently employed in a "green job" or not). Thus, we want to uncover those skills which are important to perform green tasks. Formally, this relationship can be represented by the following functional form f:

$$greenness = f(skills) + \epsilon, \tag{1}$$

where ϵ is a random error term with mean zero and independent of skills. Moreover, the variable *greenness* contains the relative number of green tasks $T_G/(T_G+T_N)$ of an occupation and the term *skills* contains a vector of skills. The problem is that the functional form f is unknown. Hence, we estimate it by using observed data from O*NET. In principle, there exists an infinite number of possible estimates of f, which we call $\hat{f}$. The aim is to find a suitable functional form with which the greenness of an occupation can be predicted accurately, i.e., yielding a low prediction error $\hat{e} = greenness - \hat{f}(skills)$.

To find a suitable functional form, we train four widely used models. Each of them imposes different restrictions to learn $\hat{f}$. In particular, we use a simple Ordinary Least Squares (OLS) regression, the Least Absolute Shrinkage and Selection Operator (LASSO) and the Ridge regression. These three models are parametric models in the sense that they make an assumption about the functional form. Moreover, we use a Random Forest regression as a representative of non-parametric models; this model has good prediction properties even when trained on data with few observations (Gunduz and Fokou, 2015). In contrast to parametric models, the functional form is not explicitly given but rather learned throughout the estimation process.

In the following section, we present the four chosen algorithms and their properties. Afterwards, we compare the prediction quality of these algorithms with the "Green General Skills" (GGS) of Vona et al. (2018). We choose the model with the best prediction quality which allows us to determine the green potential of SOC occupations.

The Four Algorithms and Their Properties

Let us start with our first algorithm, the conventional OLS regression. The OLS model assumes a linear relationship between skills and the greenness of an occupation. By minimizing the sum of squared errors between the observed and the predicted greenness of the training data,

$$\underset{\beta^{OLS}}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(greenness_i - \beta_0 - \sum_{s=1}^{p} skill_{i,s} \beta_s^{OLS} \Big)^2,$$

where $skill_{i,s}$ contains the value of the skill s of the ith observation of the training data, one obtains the famous OLS estimator:

$$\hat{\beta}^{OLS} = (X'X)^{-1}(X'Y). \tag{2}$$

The OLS estimator is composed of the matrix of skill values, X, and the vector of greenness values, Y. Since in our case, the number of explanatory variables (i.e., the number of skills in O*NET) is rather large compared to the number of observations, the OLS model may lead to overfitting. Therefore, the OLS model may have problems predicting the green potential of occupations that have not been used for training. Or in other words, the model may not generalize well to new observations. In contrast, the three other algorithms are supposed to deal well with potential overfitting (Hastie et al., 2009).[7]

In order to reduce overfitting to the training data and, thus, to increase the prediction accuracy, we use, as our second algorithm, the Ridge regression. It differs from OLS by shrinking some coefficients towards zero. It adds a penalty term to the OLS model that consists of the sum of the squared coefficients multiplied by an endogenously determined penalty parameter. Formally, the Ridge regression coefficients are obtained by minimizing the following expression (see, for example Hastie et al., 2009):

[7] There exists many more algorithms dealing with the problem of overfitting to the training data. The Ridge regression, the LASSO and the Random Forest regression we use in our analysis are three well known and often used algorithms.

$$\underset{\beta^{Ridge}}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(greenness_i - \beta_0 - \sum_{s=1}^{p} skill_{i,s} \beta_s^{Ridge} \Big)^2 + \lambda^{Ridge} \sum_{s=1}^{p} (\beta_s^{Ridge})^2.$$

This leads to the Ridge coefficients as

$$\hat{\beta}^{Ridge} = (X'X + \lambda^{Ridge} I)^{-1}(X'Y), \tag{3}$$

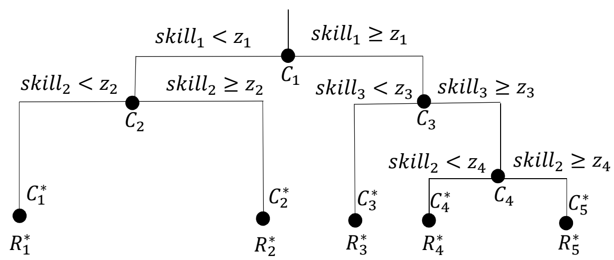
with I as the identity matrix and $\lambda^{Ridge} \geq 0$ as the endogenously chosen penalty parameter that determines the degree of shrinkage of the coefficients. The larger the λ^{Ridge}, the stronger the shrinkage of the coefficients. Note that the Ridge regression converges to the OLS model if $\lambda^{Ridge} = 0$. Compared to OLS, it is now necessary to "tune" λ^{Ridge} in order to gain a high predictive power of the Ridge regression on data not used for training. In doing so, we apply a so-called k-fold cross-validation. In particular, we follow common recommendations and apply a tenfold cross-validation (Hastie et al., 2009). This means, we randomly split our training data in ten similar sized sets, choose a value for λ^{Ridge} and train our model on nine of the ten sets. By making predictions on the particular remaining data, the mean squared error (MSE) can be calculated for all the nine sets. The average of all the resulting mean squared errors then gives the overall mean squared error. These steps are repeated for several different values of the penalty parameter, λ^{Ridge}. Ultimately, we adopt the parameter λ^{Ridge} which leads to the lowest overall mean squared error in the cross-validation exercise.

We also train a LASSO model as our third algorithm. The difference between the LASSO and the Ridge regression is the functional form of the penalty term. In case of the LASSO model, it is of linear–and not of quadratic–order. Formally, estimates of the LASSO coefficients are obtained when minimizing

$$\underset{\beta^{LASSO}}{\operatorname{argmin}} \sum_{i=1}^{N} \Big(greenness_i - \beta_0 - \sum_{s=1}^{p} skill_{i,s} \beta_s^{LASSO} \Big)^2 + \lambda^{LASSO} \sum_{s=1}^{p} |\beta_s^{LASSO}|,$$

with $\lambda^{LASSO} \geq 0$ as the endogenously chosen penalty parameter that determines the degree of shrinkage of the coefficients. However, unlike the Ridge regression, $\hat{\beta}^{LASSO}$ has no closed form solution. Moreover, compared to the Ridge regression, the LASSO model not only shrinks coefficients towards zero, but also sets some coefficients exactly to zero. To find the optimal penalty term λ^{LASSO}, we use the same steps as for the Ridge regression and choose the one with the lowest overall error in the cross-validation exercise.

Figure 2: Illustration of a Regression Tree



As our fourth algorithm, we train a Random Forest regression as proposed by Breiman (2001). In contrast to the three previous models, the functional form of the parameters is no longer linear. Instead, it can take any (non-linear) form and is learned from the data. Moreover, in contrast to other algorithms with endogenously learned functional forms, such as neural networks, it can be trained with a low number of observations. Since training a Random Forest regression requires several steps, we refer to Breiman (2001) for a rigorous formal treatment and explain only those steps which are, from our point of view, important to understand the main concept.

Figure 2 shows an example of a tree-based model. The general idea of tree-based regression models, which a Random Forest regression belongs to, is to create a hierarchy of nodes, C. At each node, C, based on the value of a particular skill, s, the predictor space is divided into two non-overlapping regions, R_j, with $j \in 1, 2$. This is done until each branch reaches a terminal node, C^*. Due to the non-overlapping splits at each previous node, all terminal nodes together divide the predictor space into several non-overlapping regions, R_j^*. Finally, the prediction value associated with a region is obtained by taking the mean value of the greenness of all training observations belonging to the particular region.

To find the optimal splitting criterion at each node, C, a top-down approach is used (for details see, for example, Hastie et al., 2009). Starting at the first node, C_1, one calculates the sum of squared prediction errors for all possible skills, s, and possible cutpoints, z,[8]

[8]Note that all values of a skill, s, which are part of the training data define the set of possible cut points, z.

$$error_C = \sum_{C:x_i \in R_1(s,z)} (greenness_i - \widehat{greenness}_{R_1})^2 + \sum_{C:x_i \in R_2(s,z)} (greenness_i - \widehat{greenness}_{R_2})^2,$$

where $\widehat{greenness}_{R1}$ is the mean of the greenness value of all training observations belonging to the first region $R_1(s, z)$, and $\widehat{greenness}_{R_2}$ the mean value belonging to the other region $R_2(s, z)$. The optimal split rule at the node, C, is given by the skill-cutpoint tupel, (s, z), which leads to the lowest error ($error_C$). In our example in Figure 2, the tupel at C_1 is $(s = skill_1, z = z_1)$. This procedure is continued until each branch reaches a terminal node, C^*. A node becomes a terminal node if either a stopping criterion is fullfilled or if it contains only one training observation.

This explains the basic idea of regression trees. However, a single tree may strongly overfit to the training data. In the extreme case, each terminal node may contain only one observation. In this case, the training data has been learned perfectly by the tree. Hence, in order to reduce the issue of overfitting to the training data, the Random Forest algorithm builds several trees. The idea behind this is that, although a particular tree may strongly overfit, the bias cancels out when a lot of trees are "grown" based on random samples. Following Breiman (2001), we introduce two kinds of randomness. First, for each tree we use a randomly drawn bootstrap sample of occupations of size 400. Second, we use at each tree-node a randomly selected subset of skills as possible split candidates. This mitigates the possibility that each tree's nodes contain similar skills s which the splitting criterion is based on.

Following general suggestions, we randomly select one third of our skills (i.e., 114/3) as candidates for splitting at each node, C. Furthermore, following the standard setup of Breiman (2001), we stop growing a branch of a tree when a node contains 5 observations. As a last point, we have to decide how many trees to grow for our Random Forest regression. Usually, the number of grown trees is increased until the Random Forest's overall mean error no longer shrinks, whereas the overall mean error is calculated by using the unused part of the training data. In our analysis, this is the case for approximately 1200 trees.

After we have grown the trees of the Random Forest, the model can be used to predict the greenness of an occupation. This works as follows. An observation is sent through all grown trees. At each node of a tree, the previously determined decision rule is applied to the corresponding skill value of the occupation. Based on this, it is decided which branch to take until a terminal node is reached. This procedure is performed for all the 1200 trees. That is, for every occupation in the data, we get 1200 predictions for its greenness. Subsequently, the final predicted value of greenness of an occupation is obtained by taking the average of all the 1200 prediction values of the reached terminal nodes across all grown trees.

Training the Algorithms and Comparing their Goodness of Prediction

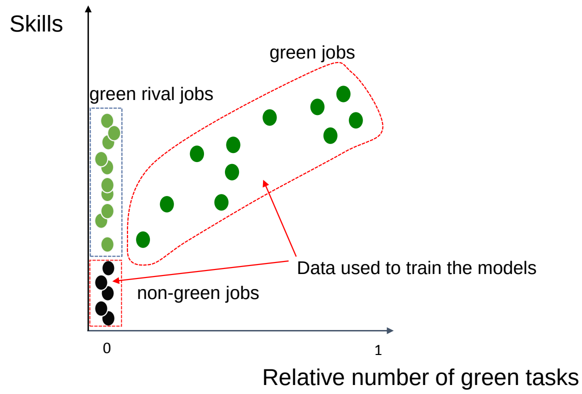
Having explained our algorithms, we show next how the O*NET data is used to train the models. As a measure of greenness of an occupation i, we use the relative number of green tasks, $greenness_i = T_G/(T_G + T_N)$. Considering only those 135 occupations with a positive number of green tasks, the median value of $greenness$ is 0.29 and the mean 0.46. Thus, even for occupations performing green tasks, the fraction of green tasks is rather low. If instead all 965 O*NET occupations are considered, the median value of $greenness$ is 0 and the mean 0.064. Further, O*NET provides valuable information on the skills required by every occupation. As explained in Section 2, O*NET states importance (IM) and level values (LV) for 114 general skills for all occupations. For our explanatory variables, $skills$, we consider both sources of information by using a weighting scheme. In particular, we weight importance by 0.7 and level by 0.3 to calculate the value of a skill s of occupation i: $skill_{i,s} = IM_{i,s}^{\alpha} LV_{i,s}^{1-\alpha}$, with $\alpha = 0.7$.[9] As a last step, we apply common practice and standardize the value of each skill by subtracting its mean and dividing it by its standard deviation (i.e., we use z-scores).[10]

We may encounter some problems when using O*NET data to identify the skills which are important to perform green tasks. The reason is that there might be occupations in O*NET that do not perform green tasks but have similar skills compared to those with a positive number of green tasks. In other words, the data suggests that skills that are important for occupations with green tasks are, at the same time, also important for occupations without green tasks. Hence, if we were to train our models using all O*NET occupations, this may likely distort the process of identifying those skills that best explain the green potential of an occupation. In order to minimize this problem, we use the following remedy. Using Bowen et al. (2018)'s classification of "green rival jobs" (see Section 1) allows us to identify those occupations among the occupations with zero green tasks that may have a completely different skill set compared to occupations with

[9] The results remain very robust when considering other values of α in the interval 0 to 1.

[10] Note that we do not standardize the response variable, i.e., the greenness remains within the value range of zero (no green tasks are performed) and one (only green tasks are performed).

Figure 3: O*NET Occupations Used to Train the Models



a positive number of green tasks. The idea is illustrated in Figure 3.

Thus, to train our models, we consider only occupations which have either (i) a positive number of green tasks (labeled in Figure 3 as "green jobs") or (ii) have zero green tasks and are very unlikely to be able to perform green tasks (labeled in Figure 3 as "non-green jobs").[11] This is the case for 304 occupations which form our group (ii). Together with the 135 green occupations from group (i), our training data consists of 439 observations. One point is worth mentioning. There are many technical occupations among those with a positive number of green tasks in O*NET. As a consequence, the skills positively associated with being able to perform green tasks could, on average, be biased towards more technical skills. This, in turn, may bias our estimation of the green potential towards more technical occupations. As the reason for this bias lies in the data, our analysis has this limitation in common with Bowen et al. (2018), Vona et al. (2018) and Consoli et al. (2016).

In order to test the goodness of prediction of our models, we randomly split our full data set of 439 occupations into a training and testing data set. We use a sample of 100 randomly drawn observations for testing and the rest for training. Thus, after having trained the four algorithms, we use our test data–i.e., the subset of data which has not been used for training–to analyze

[11] In other words, to train our models we exclude all O*NET occupations that are classified by Bowen et al. (2018) as "green rival jobs". This approach bears some similarity to Vona et al. (2018). However, since the latter are interested in identifying skills that are positively associated with the greenness of an occupation, they consider only occupations that either perform green tasks or are similar to occupations that perform green tasks. In order to identify the latter occupations, they use a different approach than Bowen et al. (2018). In particular, they take all occupations of a 3-digit SOC occupational group if at least one occupation within this 3-digit group has a positive number of green tasks. Another possibility would be to train the algorithms by using only occupations with a positive number of green tasks. However, this would reduce the sample. Moreover, such an approach would train models that discard information of skills that are negatively related with being able to perform green tasks.

the prediction goodness of our trained algorithms. In doing so, we use two common performance measures, namely the MSE of prediction and the multiclass receiver operating characteristic curve (MROC). In order to calculate these measures, we repeat the training procedure 50 times and each time calculate the prediction error of our trained models when applied to the test data. Table 1 shows the MSE of predicting the relative number of green tasks using the testing data as inputs. The Ridge regression has the lowest MSE with 0.0378 and OLS the largest with 0.0490. Hence, the relationship between skills and green tasks can be well approximated by a linear functional form.[12] In addition, as mentioned before, the LASSO sets some coefficients exactly to zero and the Ridge regression only shrinks coefficients towards zero. As a result, the smaller error term of the Ridge regression compared to the LASSO indicates that a rather large number of skills and not only a few skills are important to predict the green potential of occupations.[13]

Table 1: Prediction of the Mean Squared Error of Different Models

OLS	LASSO	Ridge	Random Forest
0.0490	0.0434	0.0378	0.0394

As pointed out in the Introduction, the approach of Vona et al. (2018) is primarily designed to identify green skills and to continuously describe occupations in terms of these selected skills. Hence, their approach could also be used to characterize the green potential of occupations (although this is not the main purpose of their analysis). To further investigate the goodness of prediction of our algorithms, we thus want to compare them to the GGS from Vona et al. (2018). However, the MSE can only be calculated if the actual and predicted value of an outcome variable is of the same metric. In Vona et al. (2018) the predicted outcome (i.e., their GGS) is a function of skills. And since the skills values are of a different metric than the greenness measure (i.e., the relative number of green tasks), it is not possible to calculate the MSE if one takes the GGS of Vona et al. (2018) as a proxy for an occupation's greenness.[14] Hence, it is not possible to use the MSE to compare the goodness of our algorithms with the indicators proposed by Vona et al. (2018).

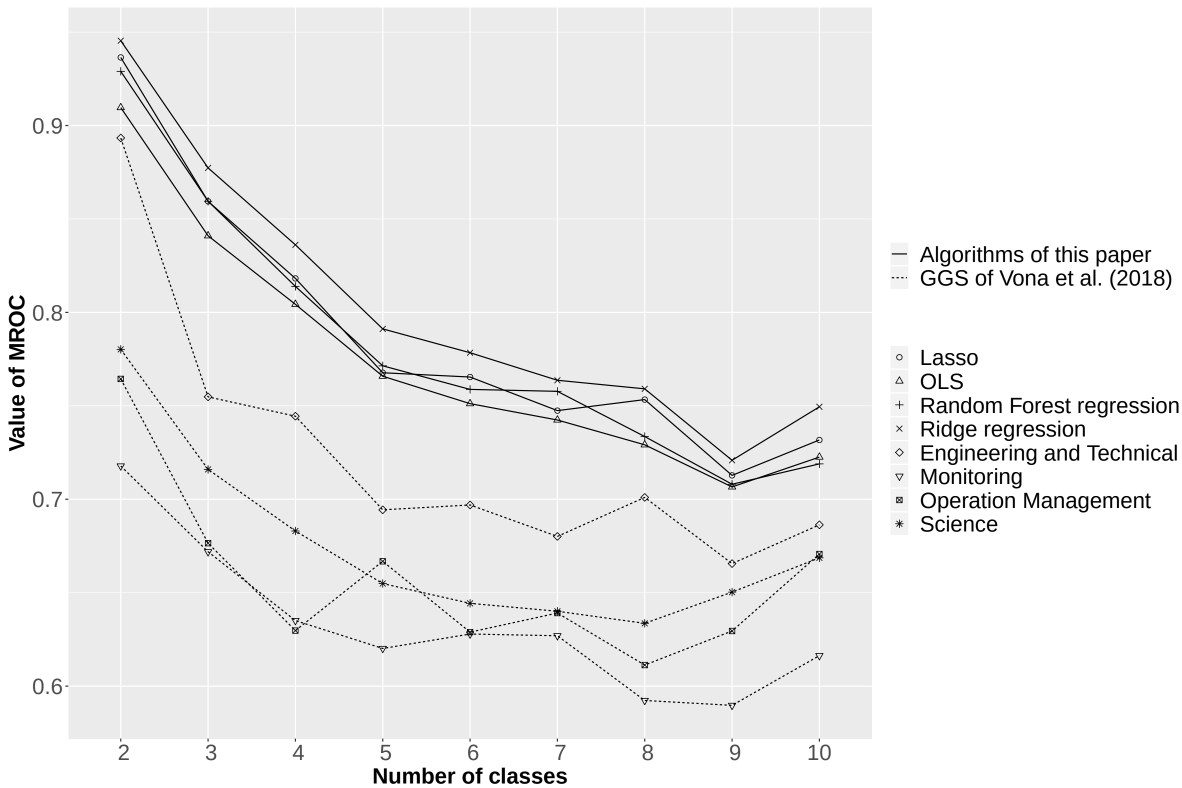
To compare the goodness of our algorithms with the one of Vona et al. (2018), we use the

[12]We have varied the number of training sessions. Moreover, we have also varied the size of the training sample. In general, these variations have had no influence on the qualitative order of the MSE of the algorithms. This leaves us confident, that for our particular task, the Ridge regression indeed outperforms the other algorithms.

[13]See Hastie et al. (2009) for a general discussion.

[14]In contrast to our approach, Vona et al. (2018)'s approach based on a PCA does not directly identify those skills which are important for green tasks. Instead, PCA bundles variables with a high correlation together to a new variable. This leads to new meta variables with zero correlation among each other.

Figure 4: Multiclass Receiver Operating Characteristic Curves



MROC. In contrast to the MSE, only the ranking between the predicted and the actual value is important for the MROC. The MROC ranks occupations according to an outcome value, in our case the greenness of an occupation. Based on this ranking, occupations are attributed to a pre-determined number of classes.[15] This allows calculating the proportion of correctly classified predictions. In doing so, the class to which an occupation is assigned to based on the predicted value is compared to the class to which an occupation is assigned to based on the actual value. The value of the MROC is then calculated by varying endogenously the class boundaries. In particular, we use a pairwise MROC as proposed by Hand and Till (2001). Intuitively, the MROC shows the probability that a randomly-drawn occupation with a higher greenness is ranked more highly than a randomly-drawn occupation with a lower greenness. Therefore, a random guess would lead to an MROC value of 0.5 and a perfect prediction method to a value of 1.

Figure 4 shows the MROC for different prediction algorithms and a different number of classes.

[15]In case of two classes, each observation belongs to one of four possible outcomes. An observation can be classified as true positive, false positive, true negative or false negative. The so-called receiver operating characteristic curve (ROC) calculates the overall share of correctly classified observations to all observations when the cut-off of belonging to the positive class increases from the lowest possible value to the largest possible value. The MROC extends the ROC to n-classes.

As one can see, the Ridge regression is slightly superior to the LASSO, to the Random Forest regression and to the OLS regression. More importantly, all our supervised learning algorithms perform better for all shown classes than the best performing measure of Vona et al. (2018), namely their GGS “Engineering and Technical”.[16] Their GGS “Science” predicts significantly worse than our trained algorithms. The other two GGS, “Operation Management” and “Monitoring”, are rather poor predictors. In the case of more than four classes, the approach by Vona et al. (2018) would be only slightly better than guessing. The relatively poor performance of their GGS to predict an occupations ability to perform green tasks may be due to two reasons. First, Vona et al. (2018)’s approach weights all ”green skills” that form a GGS identically. This means, the value of a GGS is obtained by taking the unweighted average of all skills that make up the GGS. In other words, all skills that form a GGS are implicitly assumed to be equally important. Second, their approach does not take skills into account which are *not needed* to perform green tasks. However, such skills may contain useful information to the extent that they are negatively correlated to an occupation’s potential to perform green tasks. In contrast, our supervised learning approach takes these aspects into account, leading to higher predictive accuracy.

To summarize, our proposed supervised learning algorithms predict green tasks better than if one used the proposed method to identify green skills of Vona et al. (2018) based on PCA-grouped OLS-estimates. Moreover, since the Ridge regression shows the best performance with regard to our two goodness of prediction measures, MSE and MROC, we use it in Section 4 to determine the potential of SOC occupations to perform green tasks.

Before doing so, we train the final model. It is common practice to use both, the test and training data, to determine the final model. Thus, we train a Ridge regression using *all 439* observations. The Ridge regression shrinks the coefficients of skills with low prediction power towards zero. Thus, the coefficients are not unbiased, meaning that they would not converge to the ”true” parameters. This makes it difficult to interpret single coefficients of the Ridge regression. Moreover, for the same reason, standard errors or significance levels are not very meaningful. What counts, however, is the overall goodness of prediction of the model. Thus, biased coefficients is not

[16] Vona et al. (2018)’s GGS “Engineering and Technical” results from the average of the O*NET skills “Engineering and Technology”, “Design”, “Building and Construction”, “Mechanical”, “Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment” and “Estimating the Quantifiable Characteristics of Products, Events, or Information”. GGS “Science” results from: “Biology” and “Physics”. GGS “Operation Management” results from: “System Analysis”, “Systems Evaluation”, “Updating and Using Relevant Knowledge” and “Provide Consultation and Advice to Others”. GGS “Monitoring” results from: ”‘Evaluating Information to Determine Compliance with Standards” and “Law and Government”.

Table 2: Important Coefficients of the Ridge Regression

General Skills from O*NETt	Ridge coefficients
Achievement	0.07
Building and Construction	0.15
Chemistry	0.06
Controlling Machines and Processes	-0.07
Developing and Building Teams	0.06
Economics and Accounting	0.07
Engineering and Technology	0.07
Fine Arts	-0.09
Foreign Language	-0.08
Geography	0.08
Medicine and Dentistry	-0.06
Operations Analysis	-0.07
Physics	0.08
Programming	0.06
Provide Consultation and Advice to Others	0.06
Public Safety and Security	0.07
Support	-0.13
Systems Analysis	0.06
Telecommunications	-0.06

an issue for our main analysis.

Nevertheless, it may be interesting to show the coefficients of the Ridge regression as they give an idea on which ground our empirically trained model predicts the greenness of occupations. Table 2 depicts only the coefficients of skills with a value quite different from zero.[17] In other words, it shows only those skills that are associated with a high or low predicted greenness. As one can see, technical skills are indeed associated with a high level of greenness. For example, the skill "Building and Construction" has the largest coefficient with a value of 0.15, followed by "Physics" and "Geography", both of which have a coefficient of 0.08. These results correspond in general to findings by Vona et al. (2018) and others. Interestingly, the coefficients of some skills, such as "Fine Arts" or "Medicine and Dentistry", are negative. This indicates that occupations requiring these skills intensively, may not be able to perform a relatively high number of the green tasks registered in O*NET.

[17]The complete list of coefficients of the Ridge regression can be found in Table A1 in the appendix.

4 The Green Potential of Occupations

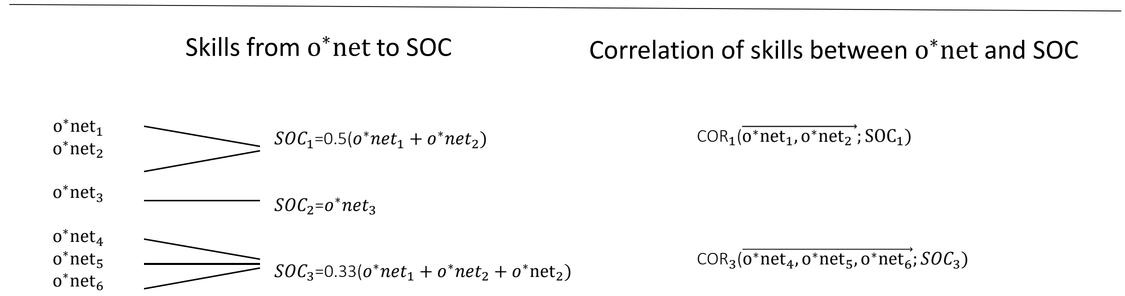
The trained model can now be used to predict the potential of an occupation to perform green tasks, as long as data for the 114 general skills are available. As occupations from O*NET are classified according to an 8-digit code and the first 6-digits correspond to the SOC-classification scheme, it is possible to transfer the skill sets from O*NET to SOC occupations by taking averages of the corresponding disaggregated occupations. The obtained values allow us to predict the potential of SOC occupations to perform green tasks. This is important because U.S. labor market data is available only at the more aggregated 6-digit SOC level.

In order to transfer skills from O*NET to SOC, we follow the previous literature (Vona et al., 2018; Bowen et al., 2018; Consoli et al., 2016) and take simple averages across all 8-digit O*NET occupations belonging to a 6-digit SOC occupation. This is exemplified in Figure 5. As one can see, the skills from the hypothetical occupations o^*net_1 and o^*net_2 result in the skills of the hypothetical occupation SOC_1. To analyze whether the skills of multiple matches are approximately similar, we calculate the correlation between the skill values of O*NET occupations belonging to the same SOC occupation and the derived skill values of that SOC occupation.[18] This is exemplified in the second column of Figure 5. In this column, COR_1 stands for the correlation between the skill values of o^*net_1 and o^*net_2 with the derived skills of the corresponding occupation SOC_1, where $SOC_1 = \frac{o^*net_1 + o^*net_2}{2}$. In our data, the correlation of skill values of O*NET and the derived ones for SOC is, on average, 0.99 and the median is also 0.99. The largest correlation is exactly one, because there are several 1:1 matches between 8-digit O*NET and 6-digit SOC codes. The lowest correlation is 0.79. Thus, due to the high correlations, using simple averages to transfer skills from O*NET to SOC seems to be a valid approach. By using the obtained skills as an input to the trained Ridge regression, we can predict each SOC occupation's potential to perform green tasks.

Note, however, that the predictions based on the Ridge regression are quite difficult to use for our subsequent analysis. There are two reasons. First, as the Ridge regression is a linear estimation technique, there exist negative green potential predictions for some occupations. Negative Ridge predictions imply that the corresponding occupations mainly possess skills which are very far from being similar to the skills of occupations that perform a relatively large number of green tasks (see Figure 1 for the mapping of skills, tasks and occupations). However, it is not very intuitive that

[18] An alternative would be to compare the variance of skill values of O*NET occupations belonging to the same SOC occupation.

Figure 5: Skills from O*NET to SOC



some occupations can have a negative green potential. Since "potential" is a non-negative value, the lowest possible green potential value should be zero. It may thus appear attractive to set all non-positive predictions to zero. However, in what follows, we intend to compare the ranking of and distance among occupations with respect to their green potential. Setting non-positive values to zero would bias this ranking.

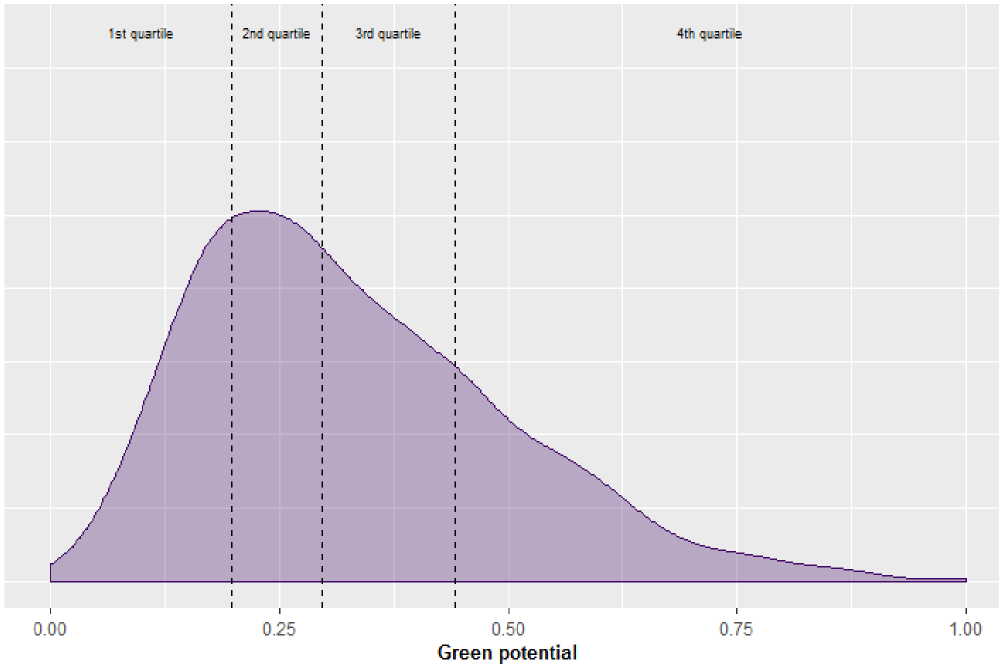
Besides negative values, there is a second challenge, which concerns the interpretation of differences between prediction values. If we used only raw Ridge regression predictions, we could not interpret differences between occupations in a meaningful way. This is quite unsatisfying. For example, there would be no meaningful way to compare the difference between occupation SOC_1 with a green potential prediction of 0.3 and occupation SOC_2 with a green potential prediction of -0.08.

To solve both of the above-mentioned challenges we decided to normalize our Ridge regression predictions on a [0,1]-scale. This solves the problem of negative prediction values as it assigns a value between 0 and 1 to all observations.[19] At the same time, it becomes much easier to interpret the values. Once we have normalized our Ridge predictions, all values have the top ranked occupation as their reference point. Note that the top-ranked occupation turns out to be "Environmental Engineers". For example, if an occupation has a normalized prediction value of 0.5, this suggests that this occupation has 50% of the green potential *compared* to "Environmental Engineers".

Figure 6 plots the distribution of our normalized Ridge regression predictions, i.e., our measure of green potential. The 25th, 50th and 75th percentiles of the distribution are given in vertical

[19] The normalized prediction values for all SOC occupations can be found in Table A3 in the appendix. Their distribution is unaffected by the normalization.

Figure 6: Distribution of Green Potential



dashed lines. The distribution's mean is 0.33 and the standard deviation is 0.17. The median green potential value is 0.30, less than one third of the top-ranked occupation's green potential. Thus, the distribution of our green potential estimates is clearly skewed towards smaller values.

Figure 6 reveals the benefit of ranking occupations continuously rather than in discrete groups. Take, for example, the three engineering occupations ”17-2081 Environmental Engineers”, ”17-3026 Industrial Engineering Technicians” and ”17-2171 Petroleum Engineers”. O*NET assigns green tasks to the first two of them and, accordingly, discrete classifications classify them as green jobs. Our green potential measure, however, captures differences between the two. It assigns the maximum green potential value of 1 to Environmental Engineers, whereas Industrial Engineering Technicians are left with a green potential estimate of 0.52. Considering the third occupation–Petroleum Engineers–reveals the second important advantage of our approach. As, according to O*NET, Petroleum Engineers do not perform green tasks, their job is not considered a green occupation. However, our measure assigns a very high green potential of 0.84 to this occupation. The reason is that the skill sets of Petroleum Engineers are very similar to those of occupations performing many green tasks. Due to this similarity of skills, Petroleum Engineers can be expected to be employed in occupations that perform green tasks or to perform green tasks within their occupation rather easily.

Another interesting finding from examining the overall distribution of green potential is that, in general, technical occupations tend to be ranked at the upper end of the distribution. Looking at the Ridge coefficients in Table 2, this result comes as no surprise. Accordingly, our green potential estimates should be interpreted primarily as measures of the potential to *directly* contribute to greening the economy by, for example, making production processes more environmentally sustainable. As a consequence, we expect that occupations with strong ties to technical skills are ranked in the upper positions, while non-technical occupations are ranked last. In order to examine this hypothesis, we take a look at the ranking of occupations. Table 3 depicts the top and bottom ranked occupations with respect to their green potential.

We have already mentioned the top rank of Environmental Engineers. Apart from that, and in line with our expectation, all the other top-ranked occupations also have strong links to technical skills. In contrast, the bottom ranked occupations are rather non-technical. Even though the tasks carried out by low-ranked occupations may contribute little or nothing to environmental damage, they have very low green potential values. This is due to the fact that their skill sets are not suited to directly contribute to greening the economy. This supports the hypothesis that our identified green potential is primarily, although not entirely, driven by technical skills.

Table 3: Top and Bottom Green Potential Estimates

Occupation (6-digit SOC)	Normalized green potential
Environmental Engineers	1.00
Chemical Engineers	0.89
Hydrologists	0.89
Civil Engineers	0.86
Agricultural Engineers	0.85
Mining and Geological Engineers, Including Mining Safety Engineers	0.85
Petroleum Engineers	0.84
Construction Managers	0.83
Architectural and Engineering Managers	0.79
Geoscientists, Except Hydrologists and Geographers	0.79
...	...
Recreational Therapists	0.06
Speech-Language Pathologists	0.06
Nursing Assistants	0.06
Actors	0.05
Orderlies	0.04
Medical Transcriptionists	0.04
Flight Attendants	0.04
Health Care Social Workers	0.02
Medical Secretaries	0.02
Art, Drama, and Music Teachers, Postsecondary	0.00

Especially from a policy perspective (e.g., for efficient targeting of labor market programs as discussed by Vona et al., 2018), it may be useful to group occupations with similar green potential values. This, for example, allows for analyzing differences with respect to education, wage rates and the age structure between green potential groups. Since our approach leads to continuous prediction values of the occupations' green potential, we could split our data into numerous groups. This would make it possible to obtain more homogeneous classes which are more accurate than, for example, the groups defined in Bowen et al. (2018).

In order to illustrate this idea, we define five green potential groups using thresholds at 0.2, 0.4, 0.6, 0.8. Keeping in mind our normalized prediction values, we can interpret these groups as green potential estimates below 20%, 40%, 60% and 80% of the top-ranked potential estimate, respectively. We then take all skills that are important to perform green tasks, i.e., those shown in Table 2 with a positive coefficient. Finally, we calculate the variance for each green potential group and each skill value. We do this also for the three groups from Bowen et al. (2018). Table 4 summarizes the results.[20] As one can see, the median and the mean of the calculated variances is almost always lower for each of our five created groups than for the three groups derived from Bowen et al. (2018). This indicates that the heterogeneity within the groups can be reduced substantially by increasing the number of groups. This, in turn, supports the claim that the green potential of occupations should be estimated on a continuous scale (Bowen et al., 2018). In this case, it is easy to attribute occupations to n different groups, where n can be any number between 1 and the number of different prediction values in the sample.

Table 4: Summary Statistic of all Calculated Variances of Skills for Different Groups

	our groups					Groups derived from Bowen et al. (2018)		
	[0.0-0.2]	[0.2-0.4]	[0.4-0.6]	[0.6-0.8]	[0.8-1.0]	non-green	pot. & indirect green	green
median	0.019	0.024	0.033	0.020	0.008	0.026	0.030	0.034
mean	0.020	0.027	0.033	0.026	0.017	0.027	0.035	0.034

Our analysis of the green potential at the occupational level leads to the following conclusions: First, consistent with the preceding literature, we find that the distribution of occupations with respect to the green potential is skewed towards lower levels of green potential. Second, the continuous measure of green potential copes well with the substantial skill heterogeneity of discrete occupational groups developed in previous studies. Third, it is interesting to note that our analysis

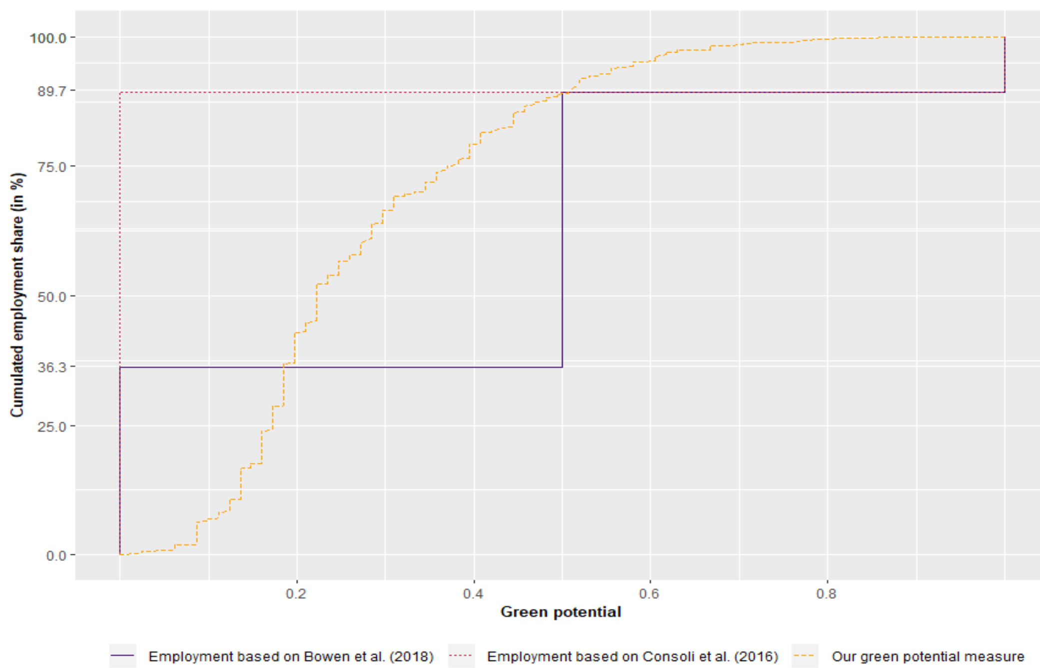
[20] Note that, building on Bowen et al. (2018), we use their "Directly Green" group as "green occupations", their "Green Rival" and "Green Increased Demand" group as potentially green occupations and their "Other" group as non-green occupations in Table 4. The results for each skill and group is found in Table A2 in the Appendix.

reveals occupations that do not currently perform any green tasks, but nevertheless have a high green potential. Finally, occupations with many technical skills tend to have a higher potential to perform green tasks, while social, cultural or artistic skills are negatively correlated with the green potential.

5 Green Potential of the U.S. Labor Market

We now take our green potential measure to the U.S. labor market data. Specifically, we illustrate the employment distribution with respect to green potential and compare it to employment estimates from the previous literature. In order to do so, we have merged our green potential estimates with occupational employment data from the Bureau of Labor Statistics.[21]

Figure 7: Green Potential and the U.S. Labor Market



Note: Employment shares for the classifications developed by Consoli et al. (2016) and Bowen et al. (2018) are taken from Bowen et al. (2018) based on 2014 employment figures. Employment shares for green potential are based on more recent occupation-level employment estimates from the Bureau of Labor Statistics for the year 2018.

Our results are shown in Figure 7, which also includes employment shares based on discrete classifications from the previous literature. The dashed curve colored in orange presents the cu-

[21] Employment data is based on estimates for the year 2018. Among the 808 6-digit SOC occupations registered in the BLS data, there have been 34 occupations where we did not have a corresponding match in the O*NET data. Therefore, we have approximated green potential values for these occupations by taking simple averages of all O*NET occupations belonging to the same 5-digit (in 14 cases) or 4-digit SOC group (in 20 cases).

mulative employment distribution in terms of our green potential measure. The red dotted curve represents a classification that differentiates between green jobs that perform green tasks (10.3% of total employment), and non-green jobs, that do not (89.7%). The solid purple curve additionally divides non-green jobs into a category of "potentially green jobs" (53.4%) and "other jobs" (36.3%).[22]

As one can quickly see, the continuous approach, represented by the dashed orange curve, allows the employment distribution to be more closely investigated and contains more information. An example helps to illustrate the reason. Let us take the two occupations "43-5041 Meter Readers, Utilities" and "51-3021 Butchers and Meat Cutters". In the discrete case, both occupations are classified as non-green but potentially green jobs. Thus, discrete classifications do not differentiate between the two. That is, a relative employment increase for Butchers and Meat Cutters would lead to an identical relative increase in "aggregate potentially green employment" as would a corresponding increase for Meter Readers. Moreover, the two could also offset each other. If, for example, relative employment for Butchers and Meat Cutters were to decrease by the same amount as relative employment increases for Meter Readers, no aggregate changes of potentially green employment would be observed.

The continuous approach, however, differentiates between the two occupations. Our own measure assigns Butchers and Meat Cutters a green potential value of 0.28, which is slightly below the median. Meter Readers show a higher value. Their green potential amounts to 0.48 which is more than twice the green potential of Butchers and Meat Cutters. Thus, a relative employment increase for Meter Readers would lead to an upward shift of the dashed orange curve at the value of 0.48. In contrast, a corresponding relative employment decrease of Butchers and Meat Cutters would be visible as a downward shift at the value of 0.28. Both changes would thus also be observed on the aggregate level. Additionally, our measure also allows us to observe where in the green potential distribution such relative employment changes occur.

Turning back to Figure 7, the dashed orange curve confirms that a vast amount of employment is concentrated at rather low green potential values. This is in line with the occupational distribution in terms of green potential plotted in Figure 6. Accordingly, the dashed orange curve in Figure 7 is very steep and, as a result, most of the U.S. employment incorporates a rather low level of green

[22]To be more specific, green jobs are defined as jobs belonging to the O*NET groups "green enhanced skills" and "green new and emerging". Potentially green jobs are either defined as "green rival jobs" in Bowen et al. (2018) or belong to the O*NET group "green increased demand". All employment figures are taken from Bowen et al. (2018).

potential. In fact, 50% of U.S. workers are employed in occupations with green potential values below 0.23. At the same time, Figure 7 also states that, for example, 4.8% of total U.S. employment has green potential values of at least 0.6. It is in the nature of a continuous measure that any chosen threshold is of course somehow arbitrary. However, the distribution of occupations' green potential shown in Figure 6 has a median of 0.3. That is, occupations above a threshold of 0.6 have at least twice the green potential of the median occupation.

It therefore seems reasonable to assume that workers in occupations above a green potential value of 0.6 should have a substantial potential to perform green tasks. Although 4.8% may not seem a very high value at first sight, it is four times the share that previous studies have attributed to jobs that currently "emerge as a response to specific needs of the green economy" (Consoli et al., 2016). And it corresponds to almost twice the share of employment associated with the production of "green goods and services" estimated by the BLS in 2011.[23] For the federal level, our green potential measure therefore refines and confirms findings from the previous literature and suggests that there is a significant share of employment that already is, or could potentially be, involved in green tasks. This is an important result with respect to a successful transformation of the U.S. to a green economy.

In addition to the country-level analysis, Bowen et al. (2018) have also provided employment estimates for individual U.S. states. They show that the share of employment which "could be involved in green economic activities" is relatively similar across states. In terms of potentially green employment, the shares of their two groups "green rival" and "green increased demand" also reveal only relatively minor differences between states.[24] Our green potential measure tends to confirm these findings, as the employment-weighted distribution of green potential turns out to be indeed similar across states (see Figure 9 in the Appendix).

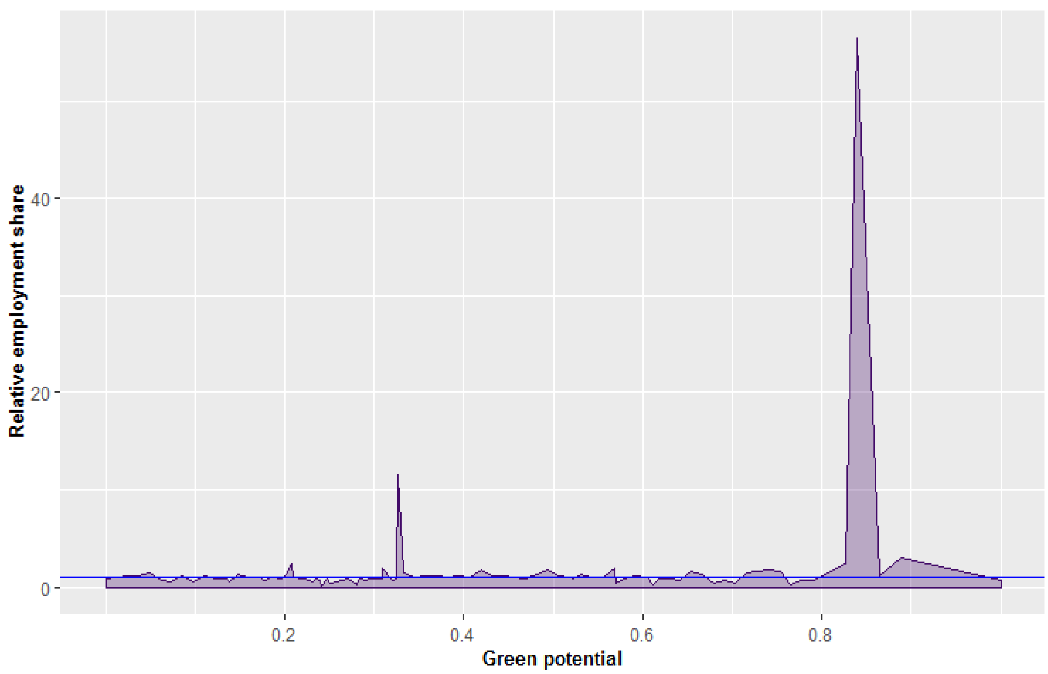
Our measure adds some interesting additional insights when comparing individual U.S. states. In what follows, we present an example that highlights the contribution of the measure in this regard. We use the states of Michigan and Texas for illustration. In Bowen et al. (2018) their respective shares of green employment are practically identical if one follows a binary classification

[23]Comparable employment estimates from previous studies (Consoli et al., 2016; Bowen et al., 2018) are based on the O*NET category "new and emerging" green occupations, which, according to Bowen et al. (2018), is closest to the BLS definition. In 2011, the BLS estimated overall green employment to 2.6% (see https://www.bls.gov/news.release/pdf/ggqcew.pdf).

[24]The mean of state level employment of the "green rival" group in Bowen et al. (2018) is 42.9% with a standard deviation of 5.2 percentage points. Adding the share of "increased demand jobs" to the share of "green rival jobs" yields a mean of 52.0% and a standard deviation of 5.2 percentage points.

(i.e., that makes a distinction between green or non-green jobs): "Directly green employment" is then estimated to 11.5% in Michigan and to 11.4% in Texas (see Bowen et al., 2018). If, additionally, "green rival jobs" are taken into account as a third group of occupations, this first finding is already somewhat refined in their analysis: In Michigan the employment share of "green rival jobs" then amounts to 41.4% and in Texas to 44.0%. Overall, however, one might conclude that the two states have very similar green employment shares, but Texas has a somewhat higher green potential.

Figure 8: Comparing U.S. States: Texas and Michigan



Note: Relative employment shares are calculated from employment estimates from the Bureau of Labor Statistics for the year 2018.

However, as shown in Figure 8, this conclusion may be misleading. For all possible values of green potential, Figure 8 plots the corresponding relative employment shares of the two states. The horizontal solid blue line indicates whether employment shares are identical (i.e., relative employment shares at a corresponding green potential value are equal to unity). A relative employment share above unity states that Texas has a larger share of employment at the corresponding value of green potential than Michigan. As Figure 8 clearly shows, Texas' main advantage in green potential arises from a single occupation, which is classified as a "green rival job" in Bowen et al. (2018). The occupation in question is "17-2171 Petroleum Engineers" which, unsurprisingly, has a much larger share of employment in Texas than in Michigan. Our measure of green potential attributes Petroleum Engineers a fairly high green potential, thus confirming that it should be counted as a potentially green occupation. However, in contrast to discrete approaches, our measure allows a more nuanced conclusion to be reached when comparing Texas and Michigan. Figure 8 suggests that Texas has minor advantages over Michigan in all occupations above a green potential of 0.8 (i.e., in those occupations with the highest green potential) and a huge advantage in terms of Petroleum Engineers. However, for the rest of the distribution above the median, the green potential values of the two states are very similar.

6 Conclusion

In this paper, we have developed an approach to measure the green potential of occupations based on a continuous scale. Our approach is based on the skills workers possess and the tasks they perform. In particular, we have trained four different machine learning algorithms, mapping skills at the level of occupations to the relative number of green tasks. The trained models allow us to predict the green potential of occupations based on their entire skill sets. We test the goodness of prediction of our trained models, whereas standard statistical tests reveal that all of our four algorithms perform significantly better in predicting the green potential of occupations than previously developed approaches. In particular, the Ridge regression showed the best performance among our algorithms.

The higher predictive accuracy of our algorithms may be due to primarily two reasons: First, we have trained a model to predict the green potential of an occupation based on the entire skill set as an input rather than on simple unweighted skill averages. Second, our approach also takes skills into account which are not needed to perform green tasks. Such skills may also contain useful information because they are negatively correlated to an occupation's potential to perform green tasks.

Our results suggest that primarily technical skills are important for a high green potential. Thus, occupations with strong ties to technical skills (e.g., engineering) are among the top-ranked occupations in terms of our green potential estimate. Our measure of green potential can therefore be interpreted as the potential of an occupation to either change production processes towards a more sustainable level, or to be able to perform new production processes which are greener compared to conventional ones.

Moreover, investigating the distribution of our green potential measure highlights that our continuous characterization of occupations can overcome two limitations of discrete approaches from the previous literature. First, it takes well into account the rather large heterogeneity in terms of the green potential of occupations within discrete classification groups used in the previous literature. We have illustrated this with two different examples. We have highlighted differences between similar engineering jobs and, in addition, we have used our continuous measure to create a larger number of discrete green potential groups. Based on this, we have shown how our approach results in a more accurate characterization of the green potential of occupations. Second, our approach ranks occupations in terms of their green potential. Thus, it enables us to compare the similarity between occupations with high green potential (but not actually involved in green tasks) with those that, according to O*NET, already perform green tasks.

Taking our green potential measure to the U.S. labor market data confirms some of the findings discovered by the preceding literature. More specifically, our measure suggests that a vast amount of aggregate employment has rather low green potential. At the same time, there is also a substantial fraction of employment that has a high potential to perform green tasks. This indicates that the U.S. labor market offers a substantial level of skills required to cope with a green transition.

Turning to the state level, our green potential measure contributes to more distinct comparisons. We have illustrated this by taking the states of Texas and Michigan as an example. Although Texas has some advantages in the upper distribution of green potential, most of the difference in green potential between the two states results from a single occupation which is Petroleum Engineers. As the skill set associated with this occupation is very similar to that of occupations which already perform green tasks, Petroleum Engineers have a very high green potential, thus favoring the green potential of Texas over that of Michigan. Therefore, our green potential measure can reveal differences between states that result from differences in occupational patterns.

For future research, it would be interesting to compare the green potential of labor markets of different economies with each other. This would reveal which labor markets are, in relative terms, well prepared for a transition towards a more sustainable economy. Transferring skills from O*NET to the internationally widely used ISCO job classification scheme and using the data in our trained algorithms may provide a promising starting point for such an analysis.

Computational Details

Our empirical results were obtained using R (version 3.6.1, see R Core Team, 2019). We have trained the algorithms presented in this paper using the packages **glmnet** (Simon et al., 2011) and **randomForest** (Liaw and Wiener, 2002). In order to calculate the MROC we used the package **pROC** (Robin et al., 2011). Furthermore, we have mainly relied on packages **dplyr** (Wickham et al., 2018) and **ggplot2** (Wickham, 2016) for data processing and data visualization.

References

Autor, D., 2013. The task approach to labor markets: an overview. Journal for Labour Market Research 46 (3), 185–199.

Bowen, A., Duffy, C., Fankhauser, S., 2016. Green growth and the new industrial revolution. Policy Brief. Grantham Research Institute on Climate Change and the Environment and Global Green Growth Institute.

Bowen, A., Kuralbayeva, K., Tipoe, E. L., 2018. Characterising green employment: The impacts of 'greening' on workforce composition. Energy Economics 72, 263–275.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32, 399.

Consoli, D., Marin, G., Marzucchi, A., Vona, F., 2016. Do green jobs differ from non-green jobs in terms of skills and human capital? Research Policy 45 (5), 1046–1060.

Dierdorff, E. C., Norton, J. J., Drewes, D. W., Kroustalis, C. M., Rivkin, D., Lewis, P., 2009. Greening of the world of work: Implications for o* net-soc and new and emerging occupations. O* NET, February.

Gunduz, N., Fokou, E., 2015. Robust classification of high dimension low sample size data. arXiv preprint arXiv:1501.00592.

Hand, D. J., Till, R. J., 2001. A simple generalisation of the area under the roc curve for multiple class classification problems. Machine Learning 45 (2), 171–186.

Hastie, T., Friedman, J., Tibshirani, R., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2 (3), 18–22.

Martinez-Fernandez, C., Hinojosa, C., Miranda, G., 2010. Greening jobs and skills labour market implications of addressing climate change oecd.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. proc: an open-source package for r and s+ to analyze and compare roc curves. BMC Bioinformatics 12, 77.

Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2011. Regularization paths for cox's proportional hazards model via coordinate descent. Journal of Statistical Software 39 (5), 1–13.

Strietska-Ilina, O., Hofmann, C., Haro, M. D., Jeon, S., 2012. Skills for green jobs: A global view. International Labour Organisation Geneva.

Vona, F., Marin, G., Consoli, D., Popp, D., 2018. Environmental regulation and green skills: An empirical exploration. Journal of the Association of Environmental and Resource Economists.

Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer.

Wickham, H., Franois, R., Henry, L., Müller, K., 2018. dplyr: A Grammar of Data Manipulation. R package version 0.7.8.

A Appendix

Tables A1 and A3 are intended to be part of an online appendix.

A.1 Tables

Table A1: Coefficients of the Ridge Regression

General Skills from O*NET	Ridge coefficients
Achievement	0.07
Active Learning	-0.01
Active Listening	-0.03
Administration and Management	0.01
Analyzing Data or Information	0.04
Assisting and Caring for Others	-0.05
Biology	0.05
Building and Construction	0.15
Chemistry	0.06
Clerical	0.01
Coaching and Developing Others	-0.01
Communicating with Persons Outside Organization	0.04
Communicating with Supervisors Peers or Subordinates	0.03
Communications and Media	0.01
Complex Problem Solving	-0.03
Computers and Electronics	-0.02
Controlling Machines and Processes	-0.07
Coordinating the Work and Activities of Others	0.05
Coordination	-0.02
Critical Thinking	-0.01
Customer and Personal Service	-0.04
Design	0.02
Developing and Building Teams	0.06
Developing Objectives and Strategies	0.01
Documenting Recording Information	0.03
Economics and Accounting	0.07
Education and Training	-0.01
Engineering and Technology	0.07
English Language	-0.02
Equipment Maintenance	0.02
Equipment Selection	-0.03
Establishing and Maintaining Interpersonal Relationships	-0.02
Estimating the Quantifiable Characteristics of Products Events or Information	0.02
Evaluating Information to Determine Compliance with Standards	0.04
Fine Arts	-0.09
Food Production	-0.03
Foreign Language	-0.08
Geography	0.08
Handling and Moving Objects	-0.04
History and Archeology	-0.02
Identifying Objects Actions and Events	-0.05
Independence	-0.03
Inspecting Equipment Structures or Material	0.01
Installation	0.05
Instructing	-0.01
Interacting With Computers	0.01
Interpreting the Meaning of Information for Others	0.02
Judging the Qualities of Things Services or People	-0.03
Judgment and Decision Making	-0.01

Law and Government	0.05
Learning Strategies	-0.03
Making Decisions and Solving Problems	-0.04
Management of Financial Resources	0.05
Management of Material Resources	0.03
Management of Personnel Resources	0.01
Mathematics	0.01
Mechanical	0.05
Medicine and Dentistry	-0.06
Monitor Processes Materials or Surroundings	0.02
Monitoring	0.01
Monitoring and Controlling Resources	0.03
Negotiation	0.05
Operating Vehicles Mechanized Devices or Equipment	0.04
Operation and Control	0.01
Operation Monitoring	0.01
Operations Analysis	-0.07
Organizing Planning and Prioritizing Work	-0.02
Performing Administrative Activities	-0.04
Performing General Physical Activities	0.02
Personnel and Human Resources	-0.03
Persuasion	0.05
Philosophy and Theology	-0.04
Physics	0.08
Production and Processing	-0.02
Programming	0.06
Provide Consultation and Advice to Others	0.06
Psychology	-0.03
Public Safety and Security	0.07
Quality Control Analysis	0.01
Reading Comprehension	0.02
Recognition	0.01
Relationships	-0.03
Repairing	0.01
Repairing and Maintaining Electronic Equipment	-0.02
Repairing and Maintaining Mechanical Equipment	-0.01
Sales and Marketing	0.01
Scheduling Work and Activities	0.01
Science	0.03
Selling or Influencing Others	0.05
Social Perceptiveness	-0.05
Sociology and Anthropology	-0.02
Speaking	-0.02
Staffing Organizational Units	-0.02
Support	-0.13
Systems Analysis	0.06
Systems Evaluation	0.05
Technology Design	-0.01
Telecommunications	-0.06
Therapy and Counseling	-0.05
Thinking Creatively	-0.02
Time Management	-0.03
Training and Teaching Others	-0.03
Transportation	0.01
Troubleshooting	0.02
Updating and Using Relevant Knowledge	-0.01
Working Conditions	0.05
Writing	0.04

Table A2: Variance of Skills within Different Green Potential Groups

	our groups					Groups of Bowen et al. (2018)		
	[0.0-0.2]	[0.2-0.4]	[0.4-0.6]	[0.6-0.8]	[0.8-1.0]	non-green	indirect green	green
Achievement	0.06	0.06	0.04	0.02	0.00	0.06	0.05	0.03
Building and Construction	0.00	0.02	0.06	0.05	0.06	0.03	0.05	0.06
Chemistry	0.02	0.03	0.03	0.04	0.05	0.03	0.04	0.04
Developing and Building Teams	0.03	0.03	0.02	0.01	0.01	0.03	0.03	0.03
Economics and Accounting	0.01	0.02	0.03	0.02	0.01	0.02	0.03	0.04
Engineering and Technology	0.01	0.02	0.04	0.04	0.01	0.02	0.05	0.05
Geography	0.02	0.02	0.03	0.04	0.02	0.02	0.03	0.04
Physics	0.01	0.02	0.03	0.04	0.02	0.02	0.04	0.04
Programming	0.00	0.02	0.03	0.02	0.02	0.01	0.02	0.01
Provide Consultation	0.03	0.03	0.02	0.01	0.00	0.03	0.03	0.02
Public Safety and Security	0.02	0.03	0.03	0.02	0.00	0.03	0.03	0.02
Systems Analysis	0.02	0.02	0.02	0.01	0.01	0.02	0.03	0.02

Table A3: Predictions from the Ridge Regression

SOC-code	Occupation	Normalized Ridge prediction
11-1011	Chief Executives	0.70
11-1021	General and Operations Managers	0.52
11-2011	Advertising and Promotions Managers	0.21
11-2021	Marketing Managers	0.44
11-2022	Sales Managers	0.49
11-2031	Public Relations and Fundraising Managers	0.36
11-3011	Administrative Services Managers	0.40
11-3021	Computer and Information Systems Managers	0.43
11-3031	Financial Managers	0.44
11-3051	Industrial Production Managers	0.70
11-3061	Purchasing Managers	0.60
11-3071	Transportation, Storage, and Distribution Managers	0.52
11-3111	Compensation and Benefits Managers	0.33
11-3121	Human Resources Managers	0.37
11-3131	Training and Development Managers	0.36
11-9013	Farmers, Ranchers, and Other Agricultural Managers	0.67
11-9021	Construction Managers	0.83
11-9031	Education Administrators, Preschool and Childcare Center/Program	0.25
11-9032	Education Administrators, Elementary and Secondary School	0.40
11-9033	Education Administrators, Postsecondary	0.33
11-9041	Architectural and Engineering Managers	0.79
11-9051	Food Service Managers	0.27
11-9061	Funeral Service Managers	0.41
11-9071	Gaming Managers	0.37
11-9081	Lodging Managers	0.35
11-9111	Medical and Health Services Managers	0.35
11-9121	Natural Sciences Managers	0.63
11-9131	Postmasters and Mail Superintendents	0.42
11-9141	Property, Real Estate, and Community Association Managers	0.60
11-9151	Social and Community Service Managers	0.28
11-9161	Emergency Management Directors	0.56
11-9199	Managers, All Other	0.62
13-1011	Agents and Business Managers of Artists, Performers, and Athletes	0.28
13-1021	Buyers and Purchasing Agents, Farm Products	0.57
13-1022	Wholesale and Retail Buyers, Except Farm Products	0.43
13-1023	Purchasing Agents, Except Wholesale, Retail, and Farm Products	0.57
13-1031	Claims Adjusters, Examiners, and Investigators	0.38
13-1032	Insurance Appraisers, Auto Damage	0.35
13-1041	Compliance Officers	0.46
13-1051	Cost Estimators	0.78
13-1071	Human Resources Specialists	0.22
13-1074	Farm Labor Contractors	0.23
13-1075	Labor Relations Specialists	0.38
13-1081	Logisticians	0.62
13-1111	Management Analysts	0.47
13-1121	Meeting, Convention, and Event Planners	0.37
13-1131	Fundraisers	0.42
13-1141	Compensation, Benefits, and Job Analysis Specialists	0.35
13-1151	Training and Development Specialists	0.19
13-1161	Market Research Analysts and Marketing Specialists	0.35
13-1199	Business Operations Specialists, All Other	0.60
13-2011	Accountants and Auditors	0.40
13-2021	Appraisers and Assessors of Real Estate	0.53
13-2031	Budget Analysts	0.37
13-2041	Credit Analysts	0.36
13-2051	Financial Analysts	0.51
13-2052	Personal Financial Advisors	0.35

13-2053	Insurance Underwriters	0.17
13-2061	Financial Examiners	0.44
13-2071	Credit Counselors	0.27
13-2072	Loan Officers	0.28
13-2081	Tax Examiners and Collectors, and Revenue Agents	0.30
13-2082	Tax Preparers	0.28
13-2099	Financial Specialists, All Other	0.53
15-1111	Computer and Information Research Scientists	0.52
15-1121	Computer Systems Analysts	0.38
15-1122	Information Security Analysts	0.46
15-1131	Computer Programmers	0.36
15-1132	Software Developers, Applications	0.44
15-1133	Software Developers, Systems Software	0.40
15-1134	Web Developers	0.31
15-1141	Database Administrators	0.35
15-1142	Network and Computer Systems Administrators	0.41
15-1143	Computer Network Architects	0.56
15-1151	Computer User Support Specialists	0.23
15-1152	Computer Network Support Specialists	0.38
15-1199	Computer Occupations, All Other	0.48
15-2011	Actuaries	0.53
15-2021	Mathematicians	0.56
15-2031	Operations Research Analysts	0.57
15-2041	Statisticians	0.43
15-2091	Mathematical Technicians	0.25
15-2099	Mathematical Science Occupations, All Other	0.25
17-1011	Architects, Except Landscape and Naval	0.77
17-1012	Landscape Architects	0.75
17-1021	Cartographers and Photogrammetrists	0.46
17-1022	Surveyors	0.67
17-2011	Aerospace Engineers	0.59
17-2021	Agricultural Engineers	0.85
17-2031	Biomedical Engineers	0.65
17-2041	Chemical Engineers	0.89
17-2051	Civil Engineers	0.86
17-2061	Computer Hardware Engineers	0.53
17-2071	Electrical Engineers	0.62
17-2072	Electronics Engineers, Except Computer	0.56
17-2081	Environmental Engineers	1.00
17-2111	Health and Safety Engineers, Except Mining Safety Engineers and Inspectors	0.75
17-2112	Industrial Engineers	0.58
17-2121	Marine Engineers and Naval Architects	0.75
17-2131	Materials Engineers	0.64
17-2141	Mechanical Engineers	0.77
17-2151	Mining and Geological Engineers, Including Mining Safety Engineers	0.85
17-2161	Nuclear Engineers	0.78
17-2171	Petroleum Engineers	0.84
17-2199	Engineers, All Other	0.72
17-3011	Architectural and Civil Drafters	0.58
17-3012	Electrical and Electronics Drafters	0.38
17-3013	Mechanical Drafters	0.54
17-3019	Drafters, All Other	0.49
17-3021	Aerospace Engineering and Operations Technicians	0.58
17-3022	Civil Engineering Technicians	0.72
17-3023	Electrical and Electronics Engineering Technicians	0.44
17-3024	Electro-Mechanical Technicians	0.49
17-3025	Environmental Engineering Technicians	0.64
17-3026	Industrial Engineering Technicians	0.52
17-3027	Mechanical Engineering Technicians	0.56
17-3029	Engineering Technicians, Except Drafters, All Other	0.59
17-3031	Surveying and Mapping Technicians	0.52
19-1011	Animal Scientists	0.67

19-1012	Food Scientists and Technologists	0.62
19-1013	Soil and Plant Scientists	0.62
19-1021	Biochemists and Biophysicists	0.56
19-1022	Microbiologists	0.47
19-1023	Zoologists and Wildlife Biologists	0.68
19-1029	Biological Scientists, All Other	0.48
19-1031	Conservation Scientists	0.60
19-1032	Foresters	0.64
19-1041	Epidemiologists	0.40
19-1042	Medical Scientists, Except Epidemiologists	0.54
19-2011	Astronomers	0.60
19-2012	Physicists	0.69
19-2021	Atmospheric and Space Scientists	0.56
19-2031	Chemists	0.56
19-2032	Materials Scientists	0.67
19-2041	Environmental Scientists and Specialists, Including Health	0.78
19-2042	Geoscientists, Except Hydrologists and Geographers	0.79
19-2043	Hydrologists	0.89
19-2099	Physical Scientists, All Other	0.63
19-3011	Economists	0.52
19-3022	Survey Researchers	0.41
19-3031	Clinical, Counseling, and School Psychologists	0.12
19-3032	Industrial-Organizational Psychologists	0.33
19-3039	Psychologists, All Other	0.15
19-3041	Sociologists	0.26
19-3051	Urban and Regional Planners	0.64
19-3091	Anthropologists and Archeologists	0.38
19-3092	Geographers	0.46
19-3093	Historians	0.30
19-3094	Political Scientists	0.33
19-3099	Social Scientists and Related Workers, All Other	0.74
19-4011	Agricultural and Food Science Technicians	0.42
19-4021	Biological Technicians	0.40
19-4031	Chemical Technicians	0.35
19-4041	Geological and Petroleum Technicians	0.53
19-4051	Nuclear Technicians	0.51
19-4061	Social Science Research Assistants	0.44
19-4091	Environmental Science and Protection Technicians, Including Health	0.72
19-4092	Forensic Science Technicians	0.43
19-4093	Forest and Conservation Technicians	0.57
19-4099	Life, Physical, and Social Science Technicians, All Other	0.48
21-1011	Substance Abuse and Behavioral Disorder Counselors	0.19
21-1012	Educational, Guidance, School, and Vocational Counselors	0.20
21-1013	Marriage and Family Therapists	0.09
21-1014	Mental Health Counselors	0.10
21-1015	Rehabilitation Counselors	0.12
21-1019	Counselors, All Other	0.14
21-1021	Child, Family, and School Social Workers	0.14
21-1022	Health Care Social Workers	0.02
21-1023	Mental Health and Substance Abuse Social Workers	0.10
21-1029	Social Workers, All Other	0.09
21-1091	Health Educators	0.22
21-1092	Probation Officers and Correctional Treatment Specialists	0.15
21-1093	Social and Human Service Assistants	0.11
21-1094	Community Health Workers	0.21
21-1099	Community and Social Service Specialists, All Other	0.17
21-2011	Clergy	0.23
21-2021	Directors, Religious Activities and Education	0.14
23-1011	Lawyers	0.41
23-1012	Judicial Law Clerks	0.22
23-1021	Administrative Law Judges, Adjudicators, and Hearing Officers	0.31
23-1022	Arbitrators, Mediators, and Conciliators	0.25

23-1023	Judges, Magistrate Judges, and Magistrates	0.26
23-2011	Paralegals and Legal Assistants	0.22
23-2091	Court Reporters	0.15
23-2093	Title Examiners, Abstractors, and Searchers	0.26
23-2099	Legal Support Workers, All Other	0.21
25-1011	Business Teachers, Postsecondary	0.31
25-1021	Computer Science Teachers, Postsecondary	0.32
25-1022	Mathematical Science Teachers, Postsecondary	0.27
25-1031	Architecture Teachers, Postsecondary	0.62
25-1032	Engineering Teachers, Postsecondary	0.72
25-1041	Agricultural Sciences Teachers, Postsecondary	0.52
25-1042	Biological Science Teachers, Postsecondary	0.36
25-1043	Forestry and Conservation Science Teachers, Postsecondary	0.60
25-1051	Atmospheric, Earth, Marine, and Space Sciences Teachers, Postsecondary	0.60
25-1052	Chemistry Teachers, Postsecondary	0.48
25-1053	Environmental Science Teachers, Postsecondary	0.49
25-1054	Physics Teachers, Postsecondary	0.51
25-1061	Anthropology and Archeology Teachers, Postsecondary	0.22
25-1062	Area, Ethnic, and Cultural Studies Teachers, Postsecondary	0.09
25-1063	Economics Teachers, Postsecondary	0.37
25-1064	Geography Teachers, Postsecondary	0.43
25-1065	Political Science Teachers, Postsecondary	0.17
25-1066	Psychology Teachers, Postsecondary	0.16
25-1067	Sociology Teachers, Postsecondary	0.16
25-1069	Social Sciences Teachers, Postsecondary, All Other	0.23
25-1071	Health Specialties Teachers, Postsecondary	0.30
25-1072	Nursing Instructors and Teachers, Postsecondary	0.20
25-1081	Education Teachers, Postsecondary	0.19
25-1082	Library Science Teachers, Postsecondary	0.27
25-1111	Criminal Justice and Law Enforcement Teachers, Postsecondary	0.33
25-1112	Law Teachers, Postsecondary	0.23
25-1113	Social Work Teachers, Postsecondary	0.16
25-1121	Art, Drama, and Music Teachers, Postsecondary	0.00
25-1122	Communications Teachers, Postsecondary	0.20
25-1123	English Language and Literature Teachers, Postsecondary	0.10
25-1124	Foreign Language and Literature Teachers, Postsecondary	0.06
25-1125	History Teachers, Postsecondary	0.09
25-1126	Philosophy and Religion Teachers, Postsecondary	0.07
25-1191	Graduate Teaching Assistants	0.20
25-1192	Home Economics Teachers, Postsecondary	0.23
25-1193	Recreation and Fitness Studies Teachers, Postsecondary	0.26
25-1194	Vocational Education Teachers, Postsecondary	0.19
25-1199	Postsecondary Teachers, All Other	0.22
25-2011	Preschool Teachers, Except Special Education	0.09
25-2012	Kindergarten Teachers, Except Special Education	0.15
25-2021	Elementary School Teachers, Except Special Education	0.17
25-2022	Middle School Teachers, Except Special and Career/Technical Education	0.28
25-2023	Career/Technical Education Teachers, Middle School	0.31
25-2031	Secondary School Teachers, Except Special and Career/Technical Education	0.20
25-2032	Career/Technical Education Teachers, Secondary School	0.47
25-2051	Special Education Teachers, Preschool	0.21
25-2052	Special Education Teachers, Kindergarten, and Elementary School	0.16
25-2053	Special Education Teachers, Middle School	0.23
25-2054	Special Education Teachers, Secondary School	0.22
25-2059	Special Education Teachers, All Other	0.22
25-3011	Adult Basic and Secondary Education and Literacy Teachers and Instructors	0.15
25-3021	Self-Enrichment Education Teachers	0.14
25-3099	Teachers and Instructors, All Other	0.22
25-4011	Archivists	0.32
25-4012	Curators	0.40
25-4013	Museum Technicians and Conservators	0.40
25-4021	Librarians	0.27

25-4031	Library Technicians	0.10
25-9011	Audio-Visual and Multimedia Collections Specialists	0.30
25-9021	Farm and Home Management Advisors	0.51
25-9031	Instructional Coordinators	0.35
25-9041	Teacher Assistants	0.09
27-1011	Art Directors	0.30
27-1012	Craft Artists	0.35
27-1013	Fine Artists, Including Painters, Sculptors, and Illustrators	0.26
27-1014	Multimedia Artists and Animators	0.25
27-1019	Artists and Related Workers, All Other	0.28
27-1021	Commercial and Industrial Designers	0.47
27-1022	Fashion Designers	0.38
27-1023	Floral Designers	0.15
27-1024	Graphic Designers	0.19
27-1025	Interior Designers	0.62
27-1026	Merchandise Displayers and Window Trimmers	0.21
27-1027	Set and Exhibit Designers	0.44
27-1029	Designers, All Other	0.35
27-2011	Actors	0.05
27-2012	Producers and Directors	0.31
27-2021	Athletes and Sports Competitors	0.33
27-2022	Coaches and Scouts	0.28
27-2023	Umpires, Referees, and Other Sports Officials	0.25
27-2031	Dancers	0.07
27-2032	Choreographers	0.10
27-2041	Music Directors and Composers	0.15
27-2042	Musicians and Singers	0.11
27-3011	Radio and Television Announcers	0.17
27-3012	Public Address System and Other Announcers	0.15
27-3021	Broadcast News Analysts	0.32
27-3022	Reporters and Correspondents	0.31
27-3031	Public Relations Specialists	0.35
27-3041	Editors	0.30
27-3042	Technical Writers	0.41
27-3043	Writers and Authors	0.25
27-3091	Interpreters and Translators	0.09
27-3099	Media and Communication Workers, All Other	0.09
27-4011	Audio and Video Equipment Technicians	0.21
27-4012	Broadcast Technicians	0.36
27-4013	Radio Operators	0.23
27-4014	Sound Engineering Technicians	0.32
27-4021	Photographers	0.19
27-4031	Camera Operators, Television, Video, and Motion Picture	0.15
27-4032	Film and Video Editors	0.22
29-1011	Chiropractors	0.27
29-1021	Dentists, General	0.37
29-1022	Oral and Maxillofacial Surgeons	0.27
29-1023	Orthodontists	0.30
29-1024	Prosthodontists	0.28
29-1029	Dentists, All Other Specialists	0.31
29-1031	Dietitians and Nutritionists	0.30
29-1041	Optometrists	0.33
29-1051	Pharmacists	0.26
29-1061	Anesthesiologists	0.25
29-1062	Family and General Practitioners	0.19
29-1063	Internists, General	0.22
29-1064	Obstetricians and Gynecologists	0.20
29-1065	Pediatricians, General	0.20
29-1066	Psychiatrists	0.12
29-1067	Surgeons	0.23
29-1069	Physicians and Surgeons, All Other	0.31
29-1071	Physician Assistants	0.21

29-1081	Podiatrists	0.23
29-1122	Occupational Therapists	0.20
29-1123	Physical Therapists	0.16
29-1124	Radiation Therapists	0.15
29-1125	Recreational Therapists	0.06
29-1126	Respiratory Therapists	0.17
29-1127	Speech-Language Pathologists	0.06
29-1128	Exercise Physiologists	0.32
29-1129	Therapists, All Other	0.14
29-1131	Veterinarians	0.30
29-1141	Registered Nurses	0.14
29-1151	Nurse Anesthetists	0.23
29-1161	Nurse Midwives	0.15
29-1171	Nurse Practitioners	0.14
29-1181	Audiologists	0.21
29-1199	Health Diagnosing and Treating Practitioners, All Other	0.16
29-2011	Medical and Clinical Laboratory Technologists	0.31
29-2012	Medical and Clinical Laboratory Technicians	0.27
29-2021	Dental Hygienists	0.10
29-2031	Cardiovascular Technologists and Technicians	0.09
29-2032	Diagnostic Medical Sonographers	0.16
29-2033	Nuclear Medicine Technologists	0.28
29-2034	Radiologic Technologists	0.17
29-2041	Emergency Medical Technicians and Paramedics	0.30
29-2051	Dietetic Technicians	0.19
29-2052	Pharmacy Technicians	0.14
29-2053	Psychiatric Technicians	0.14
29-2054	Respiratory Therapy Technicians	0.20
29-2055	Surgical Technologists	0.17
29-2056	Veterinary Technologists and Technicians	0.23
29-2057	Ophthalmic Medical Technicians	0.14
29-2061	Licensed Practical and Licensed Vocational Nurses	0.11
29-2071	Medical Records and Health Information Technicians	0.10
29-2081	Opticians, Dispensing	0.28
29-2091	Orthotists and Prosthetists	0.28
29-2092	Hearing Aid Specialists	0.21
29-2099	Health Technologists and Technicians, All Other	0.19
29-9011	Occupational Health and Safety Specialists	0.65
29-9012	Occupational Health and Safety Technicians	0.57
29-9091	Athletic Trainers	0.25
29-9092	Genetic Counselors	0.16
29-9099	Healthcare Practitioners and Technical Workers, All Other	0.21
31-1011	Home Health Aides	0.16
31-1013	Psychiatric Aides	0.10
31-1014	Nursing Assistants	0.06
31-1015	Orderlies	0.04
31-2011	Occupational Therapy Assistants	0.20
31-2012	Occupational Therapy Aides	0.10
31-2021	Physical Therapist Assistants	0.20
31-2022	Physical Therapist Aides	0.09
31-9011	Massage Therapists	0.11
31-9091	Dental Assistants	0.11
31-9092	Medical Assistants	0.17
31-9093	Medical Equipment Preparers	0.26
31-9094	Medical Transcriptionists	0.04
31-9095	Pharmacy Aides	0.16
31-9096	Veterinary Assistants and Laboratory Animal Caretakers	0.17
31-9097	Phlebotomists	0.19
31-9099	Healthcare Support Workers, All Other	0.15
33-1011	First-Line Supervisors of Correctional Officers	0.27
33-1012	First-Line Supervisors of Police and Detectives	0.38
33-1021	First-Line Supervisors of Fire Fighting and Prevention Workers	0.53

33-2011	Firefighters	0.49
33-2021	Fire Inspectors and Investigators	0.59
33-2022	Forest Fire Inspectors and Prevention Specialists	0.62
33-3011	Bailiffs	0.16
33-3012	Correctional Officers and Jailers	0.19
33-3021	Detectives and Criminal Investigators	0.37
33-3031	Fish and Game Wardens	0.46
33-3041	Parking Enforcement Workers	0.26
33-3051	Police and Sheriff's Patrol Officers	0.30
33-3052	Transit and Railroad Police	0.32
33-9011	Animal Control Workers	0.26
33-9021	Private Detectives and Investigators	0.40
33-9031	Gaming Surveillance Officers and Gaming Investigators	0.30
33-9032	Security Guards	0.20
33-9091	Crossing Guards	0.19
33-9092	Lifeguards, Ski Patrol, and Other Recreational Protective Service Workers	0.12
33-9093	Transportation Security Screeners	0.14
33-9099	Protective Service Workers, All Other	0.37
35-1011	Chefs and Head Cooks	0.33
35-1012	First-Line Supervisors of Food Preparation and Serving Workers	0.31
35-2011	Cooks, Fast Food	0.17
35-2012	Cooks, Institution and Cafeteria	0.16
35-2013	Cooks, Private Household	0.26
35-2014	Cooks, Restaurant	0.20
35-2015	Cooks, Short Order	0.14
35-2019	Cooks, All Other	0.19
35-2021	Food Preparation Workers	0.16
35-3011	Bartenders	0.25
35-3021	Combined Food Preparation and Serving Workers, Including Fast Food	0.19
35-3022	Counter Attendants, Cafeteria, Food Concession, and Coffee Shop	0.16
35-3031	Waiters and Waitresses	0.12
35-3041	Food Servers, Nonrestaurant	0.09
35-9011	Dining Room and Cafeteria Attendants and Bartender Helpers	0.16
35-9021	Dishwashers	0.26
35-9031	Hosts and Hostesses, Restaurant, Lounge, and Coffee Shop	0.17
37-1011	First-Line Supervisors of Housekeeping and Janitorial Workers	0.35
37-1012	First-Line Supervisors of Landscaping, Lawn Service, and Groundskeeping Workers	0.44
37-2011	Janitors and Cleaners, Except Maids and Housekeeping Cleaners	0.20
37-2012	Maids and Housekeeping Cleaners	0.19
37-2019	Building Cleaning Workers, All Other	0.19
37-2021	Pest Control Workers	0.44
37-3011	Landscaping and Groundskeeping Workers	0.31
37-3012	Pesticide Handlers, Sprayers, and Applicators, Vegetation	0.36
37-3013	Tree Trimmers and Pruners	0.38
37-3019	Grounds Maintenance Workers, All Other	0.36
39-1011	Gaming Supervisors	0.25
39-1012	Slot Supervisors	0.23
39-1021	First-Line Supervisors of Personal Service Workers	0.26
39-2011	Animal Trainers	0.35
39-2021	Nonfarm Animal Caretakers	0.21
39-3011	Gaming Dealers	0.17
39-3012	Gaming and Sports Book Writers and Runners	0.12
39-3019	Gaming Service Workers, All Other	0.15
39-3021	Motion Picture Projectionists	0.15
39-3031	Ushers, Lobby Attendants, and Ticket Takers	0.17
39-3091	Amusement and Recreation Attendants	0.09
39-3092	Costume Attendants	0.15
39-3093	Locker Room, Coatroom, and Dressing Room Attendants	0.09
39-3099	Entertainment Attendants and Related Workers, All Other	0.11
39-4011	Embalmers	0.22
39-4021	Funeral Attendants	0.20
39-4031	Morticians, Undertakers, and Funeral Directors	0.30

39-5011	Barbers	0.11
39-5012	Hairdressers, Hairstylists, and Cosmetologists	0.28
39-5091	Makeup Artists, Theatrical and Performance	0.30
39-5092	Manicurists and Pedicurists	0.14
39-5093	Shampooers	0.15
39-5094	Skincare Specialists	0.15
39-6011	Baggage Porters and Bellhops	0.12
39-6012	Concierges	0.15
39-7011	Tour Guides and Escorts	0.11
39-7012	Travel Guides	0.25
39-9011	Childcare Workers	0.14
39-9021	Personal Care Aides	0.09
39-9031	Fitness Trainers and Aerobics Instructors	0.09
39-9032	Recreation Workers	0.22
39-9041	Residential Advisors	0.20
41-1011	First-Line Supervisors of Retail Sales Workers	0.27
41-1012	First-Line Supervisors of Non-Retail Sales Workers	0.47
41-2011	Cashiers	0.19
41-2012	Gaming Change Persons and Booth Cashiers	0.20
41-2021	Counter and Rental Clerks	0.25
41-2022	Parts Salespersons	0.35
41-2031	Retail Salespersons	0.22
41-3011	Advertising Sales Agents	0.28
41-3021	Insurance Sales Agents	0.28
41-3031	Securities, Commodities, and Financial Services Sales Agents	0.43
41-3041	Travel Agents	0.30
41-3099	Sales Representatives, Services, All Other	0.58
41-4011	Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products	0.53
41-4012	Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products	0.41
41-9011	Demonstrators and Product Promoters	0.20
41-9012	Models	0.15
41-9021	Real Estate Brokers	0.53
41-9022	Real Estate Sales Agents	0.49
41-9031	Sales Engineers	0.57
41-9041	Telemarketers	0.16
41-9091	Door-to-Door Sales Workers, News and Street Vendors, and Related Workers	0.25
41-9099	Sales and Related Workers, All Other	0.25
43-1011	First-Line Supervisors of Office and Administrative Support Workers	0.30
43-2011	Switchboard Operators, Including Answering Service	0.07
43-2021	Telephone Operators	0.14
43-3011	Bill and Account Collectors	0.22
43-3021	Billing and Posting Clerks	0.20
43-3031	Bookkeeping, Accounting, and Auditing Clerks	0.21
43-3041	Gaming Cage Workers	0.17
43-3051	Payroll and Timekeeping Clerks	0.20
43-3061	Procurement Clerks	0.27
43-3071	Tellers	0.15
43-4011	Brokerage Clerks	0.22
43-4021	Correspondence Clerks	0.22
43-4031	Court, Municipal, and License Clerks	0.21
43-4041	Credit Authorizers, Checkers, and Clerks	0.30
43-4051	Customer Service Representatives	0.16
43-4061	Eligibility Interviewers, Government Programs	0.11
43-4071	File Clerks	0.17
43-4081	Hotel, Motel, and Resort Desk Clerks	0.25
43-4111	Interviewers, Except Eligibility and Loan	0.11
43-4121	Library Assistants, Clerical	0.16
43-4131	Loan Interviewers and Clerks	0.27
43-4141	New Accounts Clerks	0.25
43-4151	Order Clerks	0.41
43-4161	Human Resources Assistants, Except Payroll and Timekeeping	0.19
43-4171	Receptionists and Information Clerks	0.09

43-4181	Reservation and Transportation Ticket Agents and Travel Clerks	0.20
43-5011	Cargo and Freight Agents	0.36
43-5021	Couriers and Messengers	0.19
43-5031	Police, Fire, and Ambulance Dispatchers	0.16
43-5032	Dispatchers, Except Police, Fire, and Ambulance	0.41
43-5041	Meter Readers, Utilities	0.48
43-5051	Postal Service Clerks	0.28
43-5052	Postal Service Mail Carriers	0.17
43-5053	Postal Service Mail Sorters, Processors, and Processing Machine Operators	0.11
43-5061	Production, Planning, and Expediting Clerks	0.37
43-5071	Shipping, Receiving, and Traffic Clerks	0.25
43-5081	Stock Clerks and Order Fillers	0.16
43-5111	Weighers, Measurers, Checkers, and Samplers, Recordkeeping	0.20
43-6011	Executive Secretaries and Executive Administrative Assistants	0.19
43-6012	Legal Secretaries	0.16
43-6013	Medical Secretaries	0.02
43-6014	Secretaries and Administrative Assistants, Except Legal, Medical, and Executive	0.17
43-9011	Computer Operators	0.32
43-9021	Data Entry Keyers	0.23
43-9022	Word Processors and Typists	0.12
43-9031	Desktop Publishers	0.31
43-9041	Insurance Claims and Policy Processing Clerks	0.19
43-9051	Mail Clerks and Mail Machine Operators, Except Postal Service	0.19
43-9061	Office Clerks, General	0.14
43-9071	Office Machine Operators, Except Computer	0.17
43-9081	Proofreaders and Copy Markers	0.14
43-9111	Statistical Assistants	0.41
45-1011	First-Line Supervisors of Farming, Fishing, and Forestry Workers	0.51
45-2011	Agricultural Inspectors	0.44
45-2021	Animal Breeders	0.38
45-2041	Graders and Sorters, Agricultural Products	0.17
45-2091	Agricultural Equipment Operators	0.27
45-2092	Farmworkers and Laborers, Crop, Nursery, and Greenhouse	0.25
45-2093	Farmworkers, Farm, Ranch, and Aquacultural Animals	0.32
45-2099	Agricultural Workers, All Other	0.27
45-3011	Fishers and Related Fishing Workers	0.33
45-3021	Hunters and Trappers	0.42
45-4011	Forest and Conservation Workers	0.58
45-4021	Fallers	0.16
45-4022	Logging Equipment Operators	0.37
45-4023	Log Graders and Scalers	0.37
45-4029	Logging Workers, All Other	0.30
47-1011	First-Line Supervisors of Construction Trades and Extraction Workers	0.67
47-2011	Boilermakers	0.46
47-2021	Brickmasons and Blockmasons	0.51
47-2022	Stonemasons	0.54
47-2031	Carpenters	0.52
47-2041	Carpet Installers	0.35
47-2042	Floor Layers, Except Carpet, Wood, and Hard Tiles	0.42
47-2043	Floor Sanders and Finishers	0.32
47-2044	Tile and Marble Setters	0.36
47-2051	Cement Masons and Concrete Finishers	0.46
47-2053	Terrazzo Workers and Finishers	0.42
47-2061	Construction Laborers	0.44
47-2071	Paving, Surfacing, and Tamping Equipment Operators	0.36
47-2072	Pile-Driver Operators	0.42
47-2073	Operating Engineers and Other Construction Equipment Operators	0.48
47-2081	Drywall and Ceiling Tile Installers	0.44
47-2082	Tapers	0.36
47-2111	Electricians	0.56
47-2121	Glaziers	0.44
47-2131	Insulation Workers, Floor, Ceiling, and Wall	0.37

47-2132	Insulation Workers, Mechanical	0.35
47-2141	Painters, Construction and Maintenance	0.48
47-2142	Paperhangers	0.43
47-2151	Pipelayers	0.48
47-2152	Plumbers, Pipefitters, and Steamfitters	0.56
47-2161	Plasterers and Stucco Masons	0.43
47-2171	Reinforcing Iron and Rebar Workers	0.41
47-2181	Roofers	0.47
47-2211	Sheet Metal Workers	0.44
47-2221	Structural Iron and Steel Workers	0.38
47-2231	Solar Photovoltaic Installers	0.58
47-3011	Helpers–Brickmasons, Blockmasons, Stonemasons, and Tile and Marble Setters	0.37
47-3012	Helpers–Carpenters	0.46
47-3013	Helpers–Electricians	0.37
47-3014	Helpers–Painters, Paperhangers, Plasterers, and Stucco Masons	0.33
47-3015	Helpers–Pipelayers, Plumbers, Pipefitters, and Steamfitters	0.44
47-3016	Helpers–Roofers	0.40
47-3019	Helpers, Construction Trades, All Other	0.40
47-4011	Construction and Building Inspectors	0.67
47-4021	Elevator Installers and Repairers	0.57
47-4031	Fence Erectors	0.37
47-4041	Hazardous Materials Removal Workers	0.46
47-4051	Highway Maintenance Workers	0.54
47-4061	Rail-Track Laying and Maintenance Equipment Operators	0.48
47-4071	Septic Tank Servicers and Sewer Pipe Cleaners	0.47
47-4091	Segmental Pavers	0.43
47-4099	Construction and Related Workers, All Other	0.70
47-5011	Derrick Operators, Oil and Gas	0.37
47-5012	Rotary Drill Operators, Oil and Gas	0.33
47-5013	Service Unit Operators, Oil, Gas, and Mining	0.49
47-5021	Earth Drillers, Except Oil and Gas	0.42
47-5031	Explosives Workers, Ordnance Handling Experts, and Blasters	0.46
47-5041	Continuous Mining Machine Operators	0.36
47-5042	Mine Cutting and Channeling Machine Operators	0.28
47-5049	Mining Machine Operators, All Other	0.32
47-5051	Rock Splitters, Quarry	0.22
47-5061	Roof Bolters, Mining	0.36
47-5071	Roustabouts, Oil and Gas	0.28
47-5081	Helpers–Extraction Workers	0.42
49-1011	First-Line Supervisors of Mechanics, Installers, and Repairers	0.67
49-2011	Computer, Automated Teller, and Office Machine Repairers	0.31
49-2021	Radio, Cellular, and Tower Equipment Installers and Repairers	0.53
49-2022	Telecommunications Equipment Installers and Repairers, Except Line Installers	0.40
49-2091	Avionics Technicians	0.44
49-2092	Electric Motor, Power Tool, and Related Repairers	0.48
49-2093	Electrical and Electronics Installers and Repairers, Transportation Equipment	0.51
49-2094	Electrical and Electronics Repairers, Commercial and Industrial Equipment	0.62
49-2095	Electrical and Electronics Repairers, Commercial and Industrial Equipment	0.51
49-2096	Electronic Equipment Installers and Repairers, Motor Vehicles	0.44
49-2097	Electronic Home Entertainment Equipment Installers and Repairers	0.43
49-2098	Security and Fire Alarm Systems Installers	0.44
49-3011	Aircraft Mechanics and Service Technicians	0.52
49-3021	Automotive Body and Related Repairers	0.31
49-3022	Automotive Glass Installers and Repairers	0.25
49-3023	Automotive Service Technicians and Mechanics	0.46
49-3031	Bus and Truck Mechanics and Diesel Engine Specialists	0.44
49-3041	Farm Equipment Mechanics and Service Technicians	0.44
49-3042	Mobile Heavy Equipment Mechanics, Except Engines	0.54
49-3043	Rail Car Repairers	0.41
49-3051	Motorboat Mechanics and Service Technicians	0.49
49-3052	Motorcycle Mechanics	0.40
49-3053	Outdoor Power Equipment and Other Small Engine Mechanics	0.37

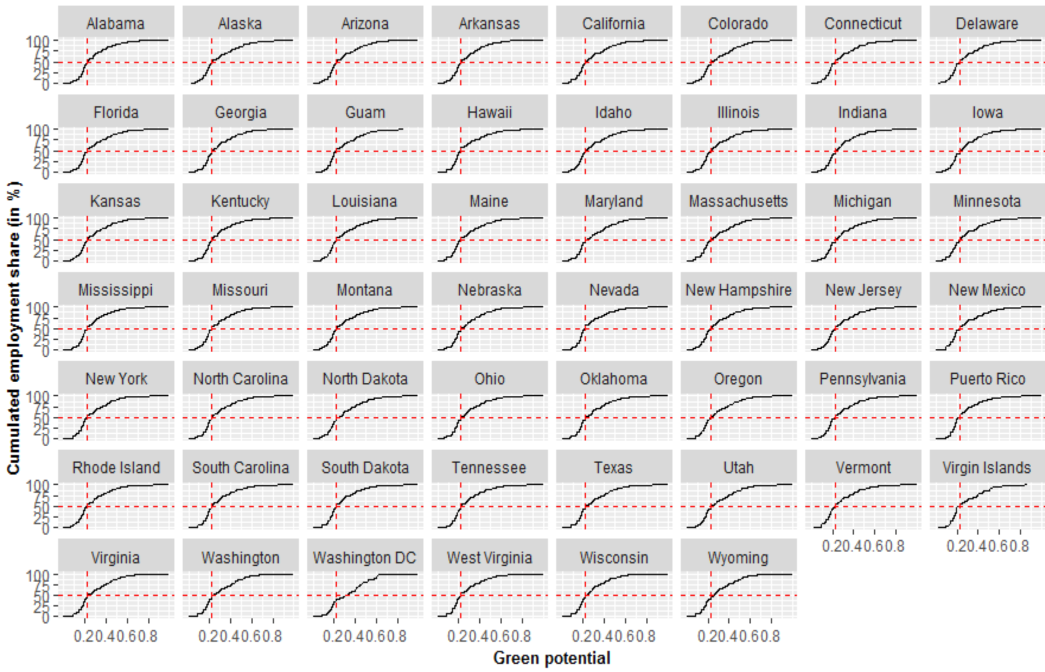
49-3091	Bicycle Repairers	0.49
49-3092	Recreational Vehicle Service Technicians	0.43
49-3093	Tire Repairers and Changers	0.40
49-9011	Mechanical Door Repairers	0.49
49-9012	Control and Valve Installers and Repairers, Except Mechanical Door	0.53
49-9021	Heating, Air Conditioning, and Refrigeration Mechanics and Installers	0.63
49-9031	Home Appliance Repairers	0.38
49-9041	Industrial Machinery Mechanics	0.46
49-9043	Maintenance Workers, Machinery	0.40
49-9044	Millwrights	0.54
49-9045	Refractory Materials Repairers, Except Brickmasons	0.27
49-9051	Electrical Power-Line Installers and Repairers	0.37
49-9052	Telecommunications Line Installers and Repairers	0.36
49-9061	Camera and Photographic Equipment Repairers	0.31
49-9062	Medical Equipment Repairers	0.44
49-9063	Musical Instrument Repairers and Tuners	0.33
49-9064	Watch Repairers	0.28
49-9069	Precision Instrument and Equipment Repairers, All Other	0.35
49-9071	Maintenance and Repair Workers, General	0.40
49-9081	Wind Turbine Service Technicians	0.57
49-9091	Coin, Vending, and Amusement Machine Servicers and Repairers	0.38
49-9092	Commercial Divers	0.54
49-9093	Fabric Menders, Except Garment	0.31
49-9094	Locksmiths and Safe Repairers	0.56
49-9095	Manufactured Building and Mobile Home Installers	0.58
49-9096	Riggers	0.37
49-9097	Signal and Track Switch Repairers	0.47
49-9098	Helpers–Installation, Maintenance, and Repair Workers	0.35
49-9099	Installation, Maintenance, and Repair Workers, All Other	0.62
51-1011	First-Line Supervisors of Production and Operating Workers	0.32
51-2011	Aircraft Structure, Surfaces, Rigging, and Systems Assemblers	0.40
51-2021	Coil Winders, Tapers, and Finishers	0.26
51-2022	Electrical and Electronic Equipment Assemblers	0.22
51-2023	Electromechanical Equipment Assemblers	0.26
51-2031	Engine and Other Machine Assemblers	0.37
51-2041	Structural Metal Fabricators and Fitters	0.33
51-2091	Fiberglass Laminators and Fabricators	0.36
51-2092	Team Assemblers	0.17
51-2093	Timing Device Assemblers and Adjusters	0.31
51-2099	Assemblers and Fabricators, All Other	0.28
51-3011	Bakers	0.19
51-3021	Butchers and Meat Cutters	0.28
51-3022	Meat, Poultry, and Fish Cutters and Trimmers	0.25
51-3023	Slaughterers and Meat Packers	0.20
51-3091	Food and Tobacco Roasting, Baking, and Drying Machine Operators and Tenders	0.26
51-3092	Food Batchmakers	0.22
51-3093	Food Cooking Machine Operators and Tenders	0.14
51-3099	Food Processing Workers, All Other	0.21
51-4011	Computer-Controlled Machine Tool Operators, Metal and Plastic	0.41
51-4012	Computer Numerically Controlled Machine Tool Programmers, Metal and Plastic	0.43
51-4021	Extruding and Drawing Machine Setters, Operators, and Tenders, Metal and Plastic	0.17
51-4022	Forging Machine Setters, Operators, and Tenders, Metal and Plastic	0.28
51-4023	Rolling Machine Setters, Operators, and Tenders, Metal and Plastic	0.27
51-4031	Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic	0.26
51-4032	Drilling and Boring Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.25
51-4033	Grinding, Lapping, Polishing, and Buffing Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.26
51-4034	Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.21
51-4035	Milling and Planing Machine Setters, Operators, and Tenders, Metal and Plastic	0.22
51-4041	Machinists	0.30
51-4051	Metal-Refining Furnace Operators and Tenders	0.22
51-4052	Pourers and Casters, Metal	0.16

51-4061	Model Makers, Metal and Plastic	0.30
51-4062	Patternmakers, Metal and Plastic	0.21
51-4071	Foundry Mold and Coremakers	0.16
51-4072	Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic	0.20
51-4081	Multiple Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.27
51-4111	Tool and Die Makers	0.30
51-4121	Welders, Cutters, Solderers, and Brazers	0.28
51-4122	Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders	0.22
51-4191	Heat Treating Equipment Setters, Operators, and Tenders, Metal and Plastic	0.25
51-4192	Layout Workers, Metal and Plastic	0.37
51-4193	Plating and Coating Machine Setters, Operators, and Tenders, Metal and Plastic	0.36
51-4194	Tool Grinders, Filers, and Sharpeners	0.26
51-4199	Metal Workers and Plastic Workers, All Other	0.31
51-5111	Prepress Technicians and Workers	0.20
51-5112	Printing Press Operators	0.27
51-5113	Print Binding and Finishing Workers	0.27
51-6011	Laundry and Dry-Cleaning Workers	0.20
51-6021	Pressers, Textile, Garment, and Related Materials	0.17
51-6031	Sewing Machine Operators	0.14
51-6041	Shoe and Leather Workers and Repairers	0.17
51-6042	Shoe Machine Operators and Tenders	0.25
51-6051	Sewers, Hand	0.11
51-6052	Tailors, Dressmakers, and Custom Sewers	0.12
51-6061	Textile Bleaching and Dyeing Machine Operators and Tenders	0.27
51-6062	Textile Cutting Machine Setters, Operators, and Tenders	0.28
51-6063	Textile Knitting and Weaving Machine Setters, Operators, and Tenders	0.09
51-6064	Textile Winding, Twisting, and Drawing Out Machine Setters, Operators, and Tenders	0.21
51-6091	Extruding and Forming Machine Setters, Operators, and Tenders, Synthetic and Glass Fibers	0.14
51-6092	Fabric and Apparel Patternmakers	0.26
51-6093	Upholsterers	0.30
51-6099	Textile, Apparel, and Furnishings Workers, All Other	0.22
51-7011	Cabinetmakers and Bench Carpenters	0.41
51-7021	Furniture Finishers	0.27
51-7031	Model Makers, Wood	0.33
51-7032	Patternmakers, Wood	0.28
51-7041	Sawing Machine Setters, Operators, and Tenders, Wood	0.20
51-7042	Woodworking Machine Setters, Operators, and Tenders, Except Sawing	0.25
51-8011	Nuclear Power Reactor Operators	0.58
51-8012	Power Distributors and Dispatchers	0.48
51-8013	Power Plant Operators	0.30
51-8021	Stationary Engineers and Boiler Operators	0.47
51-8031	Water and Wastewater Treatment Plant and System Operators	0.60
51-8091	Chemical Plant and System Operators	0.35
51-8092	Gas Plant Operators	0.48
51-8093	Petroleum Pump System Operators, Refinery Operators, and Gaugers	0.42
51-8099	Plant and System Operators, All Other	0.60
51-9011	Chemical Equipment Operators and Tenders	0.41
51-9012	Separating, Filtering, Clarifying, Precipitating, and Still Machine Setters, Operators, and Tenders	0.36
51-9021	Crushing, Grinding, and Polishing Machine Setters, Operators, and Tenders	0.37
51-9022	Grinding and Polishing Workers, Hand	0.27
51-9023	Mixing and Blending Machine Setters, Operators, and Tenders	0.38
51-9031	Cutters and Trimmers, Hand	0.17
51-9032	Cutting and Slicing Machine Setters, Operators, and Tenders	0.25
51-9041	Extruding, Forming, Pressing, and Compacting Machine Setters, Operators, and Tenders	0.23
51-9051	Furnace, Kiln, Oven, Drier, and Kettle Operators and Tenders	0.19
51-9061	Inspectors, Testers, Sorters, Samplers, and Weighers	0.25
51-9071	Jewelers and Precious Stone and Metal Workers	0.27
51-9081	Dental Laboratory Technicians	0.22
51-9082	Medical Appliance Technicians	0.36
51-9083	Ophthalmic Laboratory Technicians	0.21
51-9111	Packaging and Filling Machine Operators and Tenders	0.23
51-9121	Coating, Painting, and Spraying Machine Setters, Operators, and Tenders	0.25

51-9122	Painters, Transportation Equipment	0.27
51-9123	Painting, Coating, and Decorating Workers	0.12
51-9141	Semiconductor Processors	0.15
51-9151	Photographic Process Workers and Processing Machine Operators	0.23
51-9191	Adhesive Bonding Machine Operators and Tenders	0.23
51-9192	Cleaning, Washing, and Metal Pickling Equipment Operators and Tenders	0.22
51-9193	Cooling and Freezing Equipment Operators and Tenders	0.33
51-9194	Etchers and Engravers	0.19
51-9195	Molders, Shapers, and Casters, Except Metal and Plastic	0.30
51-9196	Paper Goods Machine Setters, Operators, and Tenders	0.22
51-9197	Tire Builders	0.17
51-9198	Helpers–Production Workers	0.19
53-1011	Aircraft Cargo Handling Supervisors	0.33
53-1021	First-Line Supervisors of Helpers, Laborers, and Material Movers, Hand	0.44
53-1031	First-Line Supervisors of Transportation and Material-Moving Machine and Vehicle Operators	0.32
53-2011	Airline Pilots, Copilots, and Flight Engineers	0.44
53-2012	Commercial Pilots	0.40
53-2021	Air Traffic Controllers	0.26
53-2022	Airfield Operations Specialists	0.44
53-2031	Flight Attendants	0.04
53-3011	Ambulance Drivers and Attendants, Except Emergency Medical Technicians	0.19
53-3021	Bus Drivers, Transit and Intercity	0.26
53-3022	Bus Drivers, School or Special Client	0.12
53-3031	Driver/Sales Workers	0.27
53-3032	Heavy and Tractor-Trailer Truck Drivers	0.36
53-3033	Light Truck or Delivery Services Drivers	0.28
53-3041	Taxi Drivers and Chauffeurs	0.27
53-4011	Locomotive Engineers	0.28
53-4012	Locomotive Firers	0.28
53-4013	Rail Yard Engineers, Dinkey Operators, and Hostlers	0.25
53-4021	Railroad Brake, Signal, and Switch Operators	0.30
53-4031	Railroad Conductors and Yardmasters	0.31
53-4041	Subway and Streetcar Operators	0.16
53-5011	Sailors and Marine Oilers	0.32
53-5021	Captains, Mates, and Pilots of Water Vessels	0.46
53-5022	Motorboat Operators	0.42
53-5031	Ship Engineers	0.53
53-6011	Bridge and Lock Tenders	0.26
53-6021	Parking Lot Attendants	0.23
53-6031	Automotive and Watercraft Service Attendants	0.36
53-6041	Traffic Technicians	0.57
53-6051	Transportation Inspectors	0.47
53-6061	Transportation Attendants, Except Flight Attendants and Baggage Porters	0.14
53-7011	Conveyor Operators and Tenders	0.20
53-7021	Crane and Tower Operators	0.35
53-7031	Dredge Operators	0.32
53-7032	Excavating and Loading Machine and Dragline Operators	0.42
53-7033	Loading Machine Operators, Underground Mining	0.27
53-7041	Hoist and Winch Operators	0.23
53-7051	Industrial Truck and Tractor Operators	0.31
53-7061	Cleaners of Vehicles and Equipment	0.23
53-7062	Laborers and Freight, Stock, and Material Movers, Hand	0.22
53-7063	Machine Feeders and Offbearers	0.22
53-7064	Packers and Packagers, Hand	0.20
53-7071	Gas Compressor and Gas Pumping Station Operators	0.48
53-7072	Pump Operators, Except Wellhead Pumpers	0.42
53-7073	Wellhead Pumpers	0.33
53-7081	Refuse and Recyclable Material Collectors	0.31
53-7111	Mine Shuttle Car Operators	0.20
53-7121	Tank Car, Truck, and Ship Loaders	0.46

A.2 Figures

Figure 9: Employment Shares in U.S. States



Note: Relative employment shares are calculated from employment estimates from the Bureau of Labor Statistics for the year 2018. The intersection of the red dotted lines highlights the green potential value (0.22) that is surpassed by 50% of total employment on the federal level. This allows comparing states relative to the more aggregate level.